PCT

WORLD INTELLECTUAL PROPERTY ORGANIZATION
International Bureau

INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(51) International Patent Classification 5 :
C12N 15/86, 15/62, 15/37
A61K 39/12

A1

(11) International Publication Number: WO 92/16636

(43) International Publication Date: 1 October 1992 (01.10.92)

(21) International Application Number: PCT/GB92/00424

(22) International Filing Date: 10 March 1992 (10.03.92)

(30) Priority data:
9105383.5 14 March 1991 (14.03.91) GB

(71) Applicant *(for all designated States except US)*: IMMUNOLOGY LIMITED [GB/GB]; 184 Cambridge Science Park, Milton Road, Cambridge CB4 4GN (GB).

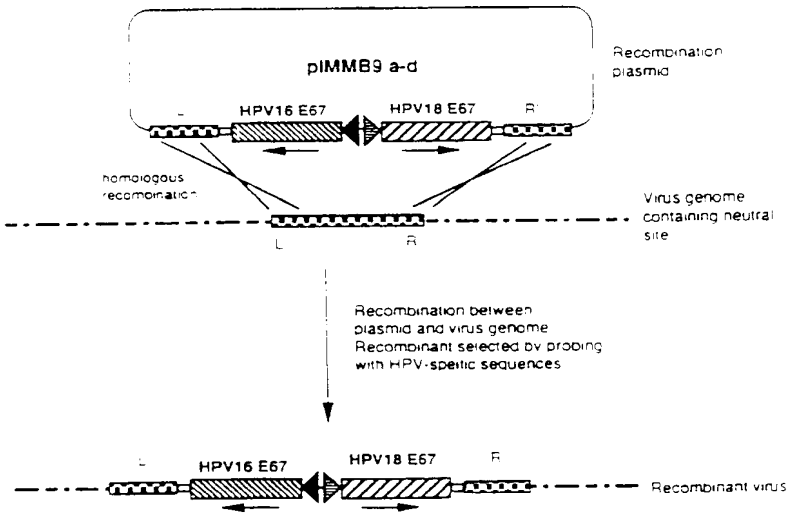
(72) Inventors; and
(75) Inventors/Applicants *(for US only)* : BOURSNELL, Michael, Edward, Griffith [GB/GB]; 46 Netherhall Way, Cambridge CB1 4NY (GB). INGLIS, Stephen, Charles [GB/GB]; 2 Rhugarve Gardens, Linton, Cambridge CB1 6LX (GB). MUNRO, Alan, James [GB/GB]; 41 Rock Road, Cambridge CB1 4UG (GB).

(74) Agent: MEWBURN ELLIS; 2 Cursitor Street, London EC4A 1BQ (GB).

(81) Designated States: AT, AT (European patent), AU, BB, BE (European patent), BF (OAPI patent), BG, BJ (OAPI patent), BR, CA, CF (OAPI patent), CG (OAPI patent), CH, CH (European patent), CI (OAPI patent), CM (OAPI patent), CS, DE, DE (European patent), DK, DK (European patent), ES, ES (European patent), FI, FR (European patent), GA (OAPI patent), GB, GB (European patent), GN (OAPI patent), GR (European patent), HU, IT (European patent), JP, KP, KR, LK, LU, LU (European patent), MC (European patent), MG, ML (OAPI patent), MN, MR (OAPI patent), MW, NL, NL (European patent), NO, PL, RO, RU, SD, SE, SE (European patent), SN (OAPI patent), TD (OAPI patent), TG (OAPI patent), US.

Published
With international search report.

(54) Title: RECOMBINANT VIRUS VECTORS ENCODING HUMAN PAPILLOMAVIRUS PROTEINS



(57) Abstract

The invention provides a recombinant virus vector for use as an immunotherapeutic or vaccine. The recombinant virus vector comprises at least one pair of nucleotide sequences heterologous to the virus and which have sufficient sequence homology that recombination between them might be expected. The pair of nucleotide sequences are arranged in the virus vector such that they are inverted with respect to each other. The virus vector is able to infect a mammalian host cell and express as polypeptide the heterologous nucleotide sequences in the host cell. For infection thought to be caused by HPV infection, the pair of nucleotide sequences encode part or all of human papillomavirus (HPV) wild-type proteins or mutant proteins immunologically cross-reactive therewith. For an immunotherapeutic or vaccine against cervical cancer, the recombinant virus vector encodes part or all of the HPV wild-type proteins HPV16E7 and HPV18E7 or mutant proteins immunologically cross-reactive therewith.

FOR THE PURPOSES OF INFORMATION ONLY

Codes used to identify States party to the PCT on the front pages of pamphlets publishing international applications under the PCT.

AT	Austria	FI	Finland	ML	Mali
AU	Australia	FR	France	MN	Mongolia
BB	Barbados	GA	Gabon	MR	Mauritania
BE	Belgium	GB	United Kingdom	MW	Malawi
BF	Burkina Faso	GN	Guinea	NL	Netherlands
BG	Bulgaria	GR	Greece	NO	Norway
BJ	Benin	HU	Hungary	PL	Poland
BR	Brazil	IE	Ireland	RO	Romania
CA	Canada	IT	Italy	RU	Russian Federation
CF	Central African Republic	JP	Japan	SD	Sudan
CG	Congo	KP	Democratic People's Republic of Korea	SE	Sweden
CH	Switzerland			SN	Senegal
CI	Côte d'Ivoire	KR	Republic of Korea	SU	Soviet Union
CM	Cameroon	LI	Liechtenstein	TD	Chad
CS	Czechoslovakia	LK	Sri Lanka	TG	Togo
DE	Germany	LU	Luxembourg	US	United States of America
DK	Denmark	MC	Monaco		
ES	Spain	MG	Madagascar		

RECOMBINANT VIRUS VECTORS ENCODING HUMAN PAPILLOMAVIRUS PROTEINS

This invention relates to recombinant virus vectors. In particular, it relates to recombinant virus vectors designed to overcome the problem of recombination between homologous nucleotide sequences. It also relates to recombinant virus vectors encoding human papillomavirus proteins; to immunotherapeutics and vaccines for conditions associated with HPV infection; to the production of a virus (e.g. vaccinia virus) engineered to express antigens encoded by human papillomavirus types 16 and 18 and to immunotherapeutics and vaccines for cervical cancer.

In recent years, strong evidence has been adduced for a link between cervical carcinoma and infection with certain types of human papillomavirus (HPV), particularly with types 16, 18, 31, 33 and 35 (Gissman et al., Cancer Cells 5,275, 1987). This is based on hybridisation studies which have indicated that more than 85-90% of biopsies from cervical tumours can be shown to contain papillomavirus DNA. HPV16 DNA is most commonly found (in about 60% of tumours) with HPV18 the next most frequent (about 20%) and the other types accounting for a further 5-10%. In many instances, tumour cells from the biopsies do not however, contain the complete genome, but rather a deleted form. The extent and location of the deleted information within the virus genome is variable, but a general feature is the retention of the part of the

genome encoding the E7 protein (Schwarz et al., Nature 314, 111, 1985). In addition, the adjacent E6-encoding region is usually present. The ubiquitous presence of the E7-encoding region in tumour cells suggests that the protein product of this gene might play a role in the induction or maintenance of the transformed phenotype. Indeed in most cell lines established from tumour biopsies, expression of the E7 gene can be detected (Smotkin & Wettstein, PNAS, 83, 4680, 1986). Furthermore, it has been shown that the E7 gene product can bind to the retinoblastoma (Rb) gene product, a recognised "anti-oncogene" in normal human cells (Munger et al., EMBO J. 8,4099, 1989). This strengthens the belief that E7 is directly involved in cell transformation.

The presence and expression of the E7 and E6 genes in tumour cells derived from cervical carcinoma biopsies, suggests the possibility that these proteins could be potential targets for the immunological recognition of the tumour cells. It is well known that viral proteins produced inside mammalian cells can be processed through a host cell pathway to short peptides, which then form a complex with host Major Histocompatibility Complex (MHC) Class 1 molecules and are transported to the cell surface. These complexes may then present a target for recognition by the host immune system. Interaction of the complex with the receptor molecule on the surface of cytotoxic T cells

(the T cell receptor) can then lead to activation of the T cells to proliferate or to destroy the recognised cell. It is possible, therefore, that the presence in the body of a population of cytotoxic T lymphocytes (CTLs) which are capable of recognising cells expressing the HPV E6 and/or E7 proteins could afford protection against the development and proliferation of cervical tumours. Indeed it has been reported that normally oncogenic mouse cells engineered to express the HPV E7 protein are unable to form tumours in mice which have been previously immunised with non-tumorigenic E7-expressing cells, and that this rejection is mediated by CD8+ lymphocytes (CTLs) (Chen et al., PNAS 88, 110, 1991). Further, the generation of an active population of such cells subsequent to tumour initiation could result in regression of the tumour.

There are numerous reports on the construction of recombinant viruses eg vaccinia viruses containing, and expressing foreign genes (Mackett & Smith, J. gen. Virol. 67,2067, 1986), and several reports of the use of these recombinant viruses to generate effective immune responses against the expressed foreign antigens. A particular advantage of this route for delivery of antigens for vaccination is that it may lead to the development of cellular as well as humoral immunity. This is because the foreign proteins will be produced inside cells of the infected individual in a

manner similar to that which occurs during natural infection. This means that they should be processed through the correct pathway to allow generation of a CTL response. In several cases, it has been demonstrated directly that immunisation with the recombinant virus is capable of producing a cellular immune response in the form of foreign antigen-specific CTLs (Moss & Flexner, Ann. Rev. Immunol., 5,305, 1987). Furthermore, vaccination of animals with recombinant vaccinia viruses expressing certain tumour-specific antigens, such as the human melanoma-associated antigen P97 (Estin et al., PNAS, 85, 1052, 1988), the bovine papillomavirus E7 protein (Meneguzzi et al., Vaccine, 8, 199, 1990) and the human breast cancer-associated antigen ETA (Hareuveni et al., PNAS, 87, 9498, 1990) has been demonstrated to result in the induction of immunity against tumour initiation and progression.

The present applicants have recognised the desirability of producing a recombinant virus vector which is useful as an immunotherapeutic or vaccine for conditions caused by HPV infection, for example for cervical cancer. With respect to cervical cancer, the art at the time of the applicants making the present invention recognised the E7 gene as having the potential to immortalise cells. Therefore, it would be felt inappropriate to incorporate the E7 gene in an immunotherapeutic. The applicants however, have

recognized the surprising usefulness of including the E7 gene in an immunotherapeutic. They have also recognized that the beneficial effects to be gained by treatment with an immunotherapeutic comprising the E7 gene are likely to outweigh by far any risk associated with the oncogenic activity of the E7 gene. Thus, an aspect of the applicants invention involves the use of a recombinant virus vector which expresses an E7 gene, as an immunotherapeutic or vaccine. Furthermore, the applicants provide embodiments of their invention in which these risks are reduced still further by specific alteration of the gene sequences in order to reduce the oncogenic potential of the E7 gene without compromising its ability to stimulate an appropriate immune response.

The present applicants have also recognized that where a number of HPV proteins which may be encoded by different HPV strains are implicated as being associated with a particular HPV-associated condition (for example, cervical carcinoma, HPV16 nd HPV18; genital warts, condyloma acuminata, respiratory papillomatosis, HPV6 and HPV11; squamous cell carcinoma in immunosuppressed individuals, HPV5 and HPV8), rather than produce a plurality of recombinant viruses engineered separately to express each of the implicated proteins, it would be advantageous to produce a single virus recombinant which is able to express part or all of the sequences of more than one

of the proteins. Thus, with respect to cervical cancer, rather than produce four recombinant viruses engineered separately to express each of the potential targets for immunological recognition of cervical tumour cells i.e. the HPV16 E6, HPV16 E7, HPV18 E6 and HPV18 E7 proteins, it would be especially advantageous to produce a single virus recombinant which is able to express the part or all of the sequences of more than one of the proteins, preferably at least two of the proteins and most preferably all four proteins. That the present applicants are able to achieve this is particularly surprising. This is because, the coding sequences for many HPV proteins are highly homologous to other equivalent HPV proteins (for example from other virus strains). Thus, the HPV16 E6 and HPV18 E6 proteins show overall homology of 62% and comprise regions of very high homology. The same is true for HPV16 E7 and HPV18 E7 which show overall homology of 57%, with particular regions of very high homology. This means that one would expect recombination to create problems such as loss of gene sequences. The applicants have, however, devised a novel strategy designed to minimise the likelihood of such recombination events and to circumvent the deleterious effect of those events should they indeed arise. Thus, surprisingly, the invention provides recombinant virus vectors which comprise at least one pair of nucleotide sequences which have sufficient sequence

homology that recombination between them might be expected. The at least one pair of nucleotide sequences may encode part or all of human papillomavirus (HPV) wild-type proteins or mutant proteins immunologically cross-reactive therewith. In particular, the invention provides a recombinant vector which can maintain stably, and express, part or all of four of the desired gene sequences from HPV16 and HPV18.

Thus, the present invention provides a recombinant virus vector for use as an immunotherapeutic or vaccine which comprises at least one pair of nucleotide sequences heterologous to said virus and which have sufficient sequence homology that recombination between them might be expected wherein said pair of nucleotide sequences are arranged in said virus vector such that they are inverted with respect to each other to reduce the likelihood of recombination events leading to loss of part or all of said sequence and said virus vector is able to infect a mammalian host cell and express as polypeptide the heterologous nucleotide sequences in said host cell. The at least one pair of nucleotide sequences may encode part or all of human papillomavirus (HPV) wild-type proteins or mutant proteins immunologically cross-reactive therewith. The pair of nucleotide sequences may encode part or all of the protein E7 from both HPV16 and HPV18 or functional equivalents

thereof. The pair of nucleotide sequences may encode part or all of the proteins E6 from both HPV16 and HPV18 or functional equivalents thereof.

The recombinant virus vector may comprise a further pair of nucleotide sequences heterologous to said virus and which (i) have sufficient sequence homology that recombination between them might be expected wherein said further pair of nucleotide sequences are arranged in said virus vector such that they are inverted with respect to each other and said virus vector is able to infect a mammalian host cell and express as polypeptide the further pair of heterologous nucleotide sequences in said host cell.

The further pair of nucleotide sequences may encode part or all of HPV wild-type proteins or mutant proteins immunologically cross-reactive therewith.

For example, the present invention also provides a recombinant virus vector which in addition to the E7 coding sequences, also comprises and is adapted to express genetic sequences encoding part or all of the protein E6 from both HPV16 and HPV18 or functional equivalents thereof. The genetic sequences may comprise sequences encoding HPV16 E6/E7 and HPV18 E6/E7 as shown in figures 1(a) and 1(b) respectively.

The genetic sequences may encode an antigenic moiety of the said proteins.

Either or both of the nucleotide sequences in a

pair of nucleotide sequences may be altered to make them less homologous than an equivalent pair of nucleotide sequences encoding wild-type HPV proteins. The alteration in nucleotide sequence may be in an area of high sequence homology. Preferably, the alteration in nucleotide sequence will not result in an alteration of the encoded amino acid sequence.

Two or more nucleotide sequences each encoding separate proteins may be fused together to form a single open reading frame. Thus the genetic sequences encoding part or all of the proteins E6 and E7 from HPV16, may be fused together to form a single open reading frame. The genetic sequences encoding part or all of the proteins E6 and E7 from HPV18, may be fused together to form a single open reading frame. The genetic sequences encoding part or all of the proteins E6 and E7 from both HPV16 and HPV18, may be fused together to form a single open reading frame. Thus, the recombinant virus vector may have the pairs of nucleotide sequences arranged according to any one of the options shown in Figure 26. Where the recombinant virus vector comprises an open reading frame having a fused genetic sequence encoding part or all of the proteins E6 and E7 from HPV16, and a separate open reading frame having fused genetic sequences encoding part or all of the proteins E6 and E7 from HPV18, the two open reading frames may be inverted with respect to one another. For example, the two open reading

frames may be arranged in the recombinant virus vector adjacent to each other. The inversion may be such that the E6 coding sequences of HPV16 and HPV18 are both located between the E7 coding sequences of HPV16 and HPV18. Alternatively, the inversion could be such that the E7 coding sequences of HPV16 and HPV18 are both located between the E6 coding sequences of HPV16 and HPV18. In particular the two open reading frames, each with its respective promoter, may be arranged next to each other in the recombinant vector. In this case the promoters may be located between the genes, which are transcribed outwardly, or the promoters may be located outside the genes, which are transcribed inwardly.

Similarly, the genetic sequences encoding part or all of the E7 protein from HPV16 and the E7 protein from HPV18 may be fused together to form a single open reading frame. The genetic sequences encoding part or all of the E6 protein from HPV16 and the E6 protein from HPV18 may be fused together to form a single open reading frame. This leads to another range of arrangements similar to those shown in Figure 26. The fusions may be via a single codon encoding a relatively small neutral amino acid e.g. glycine.

Thus the present invention also provides a recombinant virus vector which comprises a first open reading frame having a fused genetic sequence encoding part or all of the wild-type proteins E6 and E7 from

HPV16; and a separate second open reading frame having a fused genetic sequence encoding part or all of the wild-type proteins E6 and E7 from HPV18; wherein the first and second open reading frames may be inverted with respect to one another whereby either: i) the E6 coding sequences of HPV16 and HPV18 are both located between the E7 coding sequences of HPV16 and HPV18; or ii) the E7 coding sequences of HPV16 and HPV18 are both located between the E6 coding sequences of HPV16 and HPV18; and wherein any of said wild-type proteins may be replaced by a mutant protein immunologically cross-reactive therewith.

Each of the first and second open reading frames may have a corresponding promoter and the two open reading frames each with its promoter, are arranged next to each other in the virus.

The present invention also provides a recombinant virus vector wherein either: i) the promoters are located between the first and second reading frames whereby the open reading frames are transcribed outwardly; or ii) the promoters are located outside the first and second open reading frames whereby the open reading frames are transcribed inwardly.

The present invention also provides a recombinant virus vector which comprises a first open reading frame having a fused genetic sequence encoding part or all of the wild-type proteins E6 and E7 from HPV16; and a separate second open reading frame having a

fused genetic sequence encoding part or all of the wild-type proteins E6 and E7 from HPV18; wherein the E6 coding sequences of HPV16 and HPV18 are both located between the E7 coding sequences of HPV16 and HPV18; and each open reading frame has a corresponding promoter, the promoters being located between the first and second open reading frames whereby the open reading frames are transcribed outwardly; and wherein any of said wild-type proteins may be replaced by a mutant protein immunologically cross-reactive therewith.

The wild-type proteins HPV16E7 and HPV18E7 may be replaced with mutant proteins which are substantially homologous to said wild-type proteins and in which the residues cys 24 and glu 26 of wild-type protein HPV16E7 and the residues cys 27 and glu 29 of wild-type protein HPV18E7 are replaced with glycine residues.

The recombinant virus vector may be derivable from vaccinia virus.

The applicants have also recognized that for effective function as an immunotherapeutic, it is desirable for the recombinant virus to retain its ability to replicate and thereby generate an active infection in order that a cellular immune response may be mounted against the virus-encoded proteins. Thus, the applicants propose that the foreign gene sequences should be inserted into the vector virus at sites, the

disruption of which by the insertion of the heterologous gene sequences will not substantially interfere with, and therefore have a substantially adverse affect on any viral functions which relate to the replicative ability of the virus in the infected host animal. The applicants have named these sites 'neutral sites' (although the term 'neutral' should not be interpreted strictly as it is acknowledged that the disruption of these sites may have a small, but relatively speaking inconsequential adverse affect on replicative ability).

DNA sequences which affect virus replication can fall into several categories:

i) protein coding sequences;

ii) elements involved in control of gene expression; and iii) elements involved in virus DNA replication

A non-essential and neutral insertion site must therefore avoid such regions, and, such sites have been identified on the basis of nucleotide sequencing studies. Thus the genetic sequences may be inserted into neutral sites within the virus genome. One or more genetic sequences may be inserted into the same neutral site.

Neutral sites can be easily tested for according to techniques well known in the art. For example, a site may be selected, interrupted or deleted using standard methodologies and the resultant recombinant

virus placed in conditions which normally support growth of the wild type virus vector, to assess the effect of the manipulations. The pathogenicity of the virus may be further compared with that of the unmodified virus vector strain in animal models, in order to assess its level of attenuation.

In the present invention, the virus vector may be vaccinia virus. The vaccinia virus may be attenuated or disabled so that it is unable to fully replicate and esablish an extensive infection of host cells.

Vaccinia virus has been used extensively in the past for vaccination against smallpox, and its use worldwide has led to the complete eradication of the disease (Bhebehami, Microbiol. Rev., 47, 455, 1983). During the World Health Organisation (WHO) campaign to eradicate smallpox, several different strains of vaccinia virus were used as vaccines. In 1984 a meeting was sponsored by the WHO to discuss the use of vaccinia virus as live virus vectors (Bulletin of the WHO 63(3):471-477). The data in this report indicates that the number of complications associated with vaccination was lowest for the Wyeth strain of vaccinia virus, and so this strain has been chosen as a basis for the construction of the recombinant virus according to an embodiment of the present invention.

It is well known that insertion of foreign DNA into the genome of vaccinia virus at certain favoured sites, such as the thymidine kinase gene locus, can

reduce dramatically the ability of the virus to replicate in vivo. As discussed above, the aim of the therapeutic approach described here is to generate an active in vivo infection, so that a cellular immune response may be mounted against the virus encoded proteins. The present invention provides a method for inserting foreign genes at neutral sites within the genome of a virus, the disruption of which sites by the insertion will not interfere with and therefore have a substantially adverse affect on virus replication.

Where the virus is vaccinia virus, the neutral site may be identified herein within the Wyeth strain of vaccinia virus on the basis of the related WR strain nucleotide sequence. Alternatively, where other vaccinia virus strains are used, sites equivalent to those sites identified above may be used. The neutral sites may be any as identified hereinafter as A,B,C and D, or a functional equivalent.

For successful expression of foreign proteins by the recombinant virus vector the foreign genes must be placed under the control of a promoter sequence which is operable by the virus. Thus the recombinant virus vector may comprise a single promoter which controls the expression of all the heterologous genetic sequences within a single open reading frame. Alternatively, where the recombinant virus vector

encodes more than one open reading frame containing heterologous genetic sequences, the virus may comprise a first promoter which controls the expression of the genetic sequences from a first open reading frame, and one or more further promoters which control the expression of the genetic sequences from one or more further open reading frames. The promoter sequence may be virus- specific and several have been characterised so far (Davison & Moss, J. Mol. Biol., 210, 749, 1989; Davison & Moss, J. Mol. Biol., 210, 771, 1989). The single promoter and the first and one or more further promoters may be the p 7.5 promoter. There have been reports that the induction of foreign antigen-specific CTLs requires expression of the antigen early in the virus replication cycle (Coupar et al., Eur. J. Immunol., 16, 1479, 1986). Therefore, a recombinant virus as provided by the present invention may involve the use of the p7.5 promoter (Venkatesan et al., Cell, 125, 805, 1981) and/or the H6 promoter (Rosel et al., J. Virol, 60, 436, 1988), both of which are active both early and late in infection.

As mentioned earlier, it has been reported that the E7 gene on its own has the potential to immortalise cells (Phelps et al., Cell 53, 539, 1988). In an embodiment of the present invention, the strategy for expression of the protein involves production of E7 as a fusion protein with E6, which is

unlikely to retain biological function. Embodiments of the invention provide for reducing this risk still further, by making changes within the E7 gene which are known to destroy its oncogenic capacity (Chesters et al., J. Gen Virol. 71, 449. 1990). Thus in the recombinant virus vectors of the present invention, the genetic sequences encoding part or all of the E7 proteins may be altered from the equivalent wild type sequences, in order to render the sequences, used in the recombinant virus vectors less oncogenic than their equivalent wild type sequences.

The present invention also provides pharmaceuticals comprising recombinant virus vectors as herein defined. The pharmaceutical may be for use against a condition caused by HPV infection which comprises an immunotherapeutically effective amount of a recombinant virus vector. The pharmaceutical may be for use against cervical cancer.

The pharmaceutical may be a vaccine to immunise against a condition caused by HPV infection which comprises an amount of recombinant virus vector as herein provided which when administered to a recipient can specifically activate cells of the immune system to HPV proteins. The vaccine may be for immunisation against cervical cancer.

The pharmaceuticals may comprise one or more excipients. The present invention also provides methods of using the recombinant virus vectors as

herein defined to make medicaments for use as immunotherapeutics or vaccines against conditions thought to be caused by HPV infection. For example for the prophylaxis and treatment of cervical cancer.

The present invention also provides methods of treating mammalian patients with recombinant virus vectors and pharmaceuticals as herein provided.

The present invention also provides a method of determining a neutral site in a virus vector, the disruption of which by the insertion of heterologous gene sequences will not interfere with, and therefore, have a substantially adverse affect on viral function which relates to the replicative ability of the virus. The method for this determination comprises: (a) analysing a viral genome to identify open reading frames which are likely to encode functional genes, by looking for expected codon usage between spaced apart start and stop codons; and (b) selecting sites which are not in such open reading frames, likely to encode functional genes, as identified in (a) This may include selecting sites between open reading frames for sequences of functional genes and selecting sites which are in open reading frames which have some functional gene characteristics, such as an expected codon usage, but have lost other essential characteristics such as a start codon. The method may also comprise interrupting or deleting the selected sites from the viral genome and placing the resultant

virus in conditions which normally support growth of the wild type virus.

The present invention also provides neutral sites identified by use of the above methods.

The present invention provides an embodiment which shows a way of inducing a cellular immune response against the papillomavirus proteins usually expressed in cervical tumour cells by the creation of a recombinant vaccinia virus, which has been engineered to produce the HPV E6 and E7 proteins, or proteins containing HPV E6 and E7 sequences, during its replication cycle. This therapeutic vaccinia virus contains the E6 and E7 genes from both HPV16 and HPV18, the viruses most commonly associated with cervical carcinoma. Vaccination with this single virus may thus stimulate immunity to the E6 and E7 proteins of the HPV types associated with more than 80% of cervical tumours. Expression of all four gene sequences (e.g. HPV16 E6 and E7; HPV18 E6 and E7) in a single virus however presents a problem, because of the likelihood of loss of genetic sequences through recombination. The present invention provides a method for circumventing this difficulty, firstly through specific sequence alteration, in order to reduce sequence homology and secondly through their insertion into the vaccinia virus genome in such a way that if such recombination were to occur, it would not lead to loss of sequences (i.e. in inverted

orientation with respect to each other). Expression of the desired four gene sequences in the vaccinia virus genome could also be difficult (though not impossible) to achieve as independent expression units, and so the invention provides that instead, the E6 and E7 open reading frames may be fused together. A problem with standard methods for insertion of foreign information into the vaccinia virus genome is that the use of selectable markers to increase the efficiency of recombination results in the ultimate presence in the recombinant virus also of the selectable marker gene itself. Methods for insertion have been developed however, which allow subsequent elimination of these extraneous sequences (Falkner & Moss J. Virol., 64, 3108, 1990) and these are used in an embodiment of the present invention to ensure that the final recombinant vaccinia virus has only those additional sequences which are necessary for its required function.

In order that the present invention is more fully understood an embodiment will now be described in more detail with reference to the figures in which:

Figure 1(a) shows the nucleotide sequence and three-frame translation of HPV16 E6/E7 polymerase chain reaction product (underlined regions indicate the E6 and E7 coding sequences); Figure 1(b) shows the nucleotide sequence and three-frame translation of HPV18 E6/E7 polymerase chain reaction product

(underlined regions indicate the E6 and E7 coding sequences);

Figure 2 shows the cloning and modification of the HPV16 and HPV18 E6 and E7 genes;

Figure 3 shows an open reading frame plot of vaccinia virus from positions 17201-18450 of the region covered by the four fragments SalF,G,H and I; short vertical lines denote termination codons, lines topped with boxes denote initiation codons, rectangles show relevant open reading frames, and arrows show direction of upper and lower DNA strands;

Figure 4 shows an open reading frame plot of vaccinia virus from positions 21001-22000 of the region covered by the four fragments SalF,G,H and I; short vertical lines denote termination codons, lines topped with boxes denote initiation codons, rectangles show relevant open reading frames and arrows show direction of upper and lower DNA strands;

Figure 5 shows an open reading frame plot of vaccinia virus from positions 23501-25000 of the region covered by the four fragments SalF,G,H and I; short vertical lines denote termination codons, lines topped with boxes denote initiations codons, rectangles show relevant open reading frames and arrows show the direction of upper and lower strands of DNA;

Figure 6 shows a codon usage plot of vaccinia virus from positions 17201-18450 of the region covered by the four fragments SalF,G,H and I; arrows show

direction of each DNA strand;

Figure 7 shows a codon usage plot of vaccinia virus from positions 21001-22000 of the region covered by the four fragments SalF,G,H and I; arrows show direction of each DNA strand;

Figure 8 shows a codon usage plot of vaccinia virus from positions 23501-25000 of the region covered by the four fragments SalF,G,H and I; arrows show direction of each DNA strand;

Figure 9 shows the DNA sequence around site A showing translations in single letter amino acid code of genes SalF 17R and SalF 19R;

Figure 10 shows the DNA sequence around site B showing translations in single letter amino acid code of genes SalF 20R and SalF20.5R;

Figure 11 shows a comparison of the SalG2R open reading frame to the yeast guanylate kinase gene sequence;

Figure 12 shows the DNA sequence around site D showing translations in single letter amino acid code of genes HindB3R and Hind B4R;

Figure 13 shows the cloning of vaccinia virus (Wyeth strain) neutral sites;

Figure 14 shows the cloning of vaccinia virus promoter sequences;

Figure 15 shows the construction of vaccinia promoter-driven E6-7 cassette;

Figure 16 shows the cloning of the E6-7 cassette into

vaccinia virus (Wyeth strain) neutral sites;

Figure 17 is a diagram showing the recombination required to generate the final therapeutic vaccinia virus - HPV recombinant virus;

Figure 18 shows the synthetic oligonucleotides used in the construction of the therapeutic vaccinia virus HPV recombinant;

Figure 19 shows the nucleotide sequence of vaccinia virus (WR strain) from positions 17201-18450 of the region covered by the four fragments SalF,G,H and I;

Figure 20 shows the nucleotide sequence of vaccinia virus (WR strain) from positions 21001-22000 of the region covered by the four fragments SalF,G,H and I; and

Figure 21 shows the nucleotide sequence of vaccinia virus (WR strain) from positions 23501-25000 of the region covered by the four fragments SalF,G,H and I.

Figure 26 shows a variety of options for arrangment of HPV16E6 and E7 and HPV18E6 and E7 coding sequences in a recombinant virus vector.

All cloning procedures are carried out according to the protocols described in "Molecular Cloning", A Laboratory Manual, eds. Sambrook, Fritsch and Maniatis, Cold Spring Harbor Laboratory Press, 1989. All plasmids on which site directed mutagenesis is performed are of the "phagemid" type, which may be converted to single-stranded DNA by superinfection with the bacteriophage f1. Preparation and site

directed mutagenesis of single-stranded DNA, is carried out as described by Brierley et al., Cell, 57, 537, 1989. The sequence of all the synthetic oligonucleotides used are provided in figure 18.

Preparation of the E6 and E7 genes from HPV16 and HPV18 for insertion into vaccinia virus

Cloning of the HPV16 and HPV18 E6 and E7

A fragment of DNA containing the HPV16 E6/7 coding region is prepared by polymerase chain reaction (PCR) amplification from the plasmid pBR322/HPV16 (Durst et al., PNAS, 80, 3812, 1983) using the oligonucleotides SO5 and SO6. A fragment containing the same region from HPV18 is prepared by the same procedure from plasmid pBR322/HPV18 (Boshart et al., EMBO J. 3,1151) using the oligonucleotides SO1 and SO2. Plasmids pBR322/HPV16 and pBR322/HPV18 are both available from Behringwerke AG, P.O. Box 1140 D-3550, Marburg, Germany (alternatively the necessary sequences can be created synthetically from the sequence information provided by the present application).

In each case, this produces a DNA fragment of about 800 base pairs (bp) with a site for the restriction enzyme Nco 1 (CCATGG) located exactly at the beginning of the E6 gene, and a Sma1 site immediately downstream of the termination codon for the E7 gene (figure 1(a) and (b)). The products are then digested with Nco1 and Sma1, and cloned into

NcoI-SmaI digested plasmid pUC118NS (a modified version of the "phagemid" pUC118 (Viera & Messing, Methods Enzymol., 153,3, 1987) in which NcoI and SmaI sites have been created by site-directed mutagenesis within the poly-linker region) to generate the plasmid pIMS7, containing the HPV16 sequences, and pIMS8 containing the HPV18 sequences (figure 2). The use of pUC118 is not crucial to the present strategy as any plasmid which can be manipulated by site directed mutagenesis can be successfully used.

Fusion of the E6 and E7 ORFs

For insertion into vaccinia virus, the E6 and E7 genes from each HPV type, are first fused together to form a single continuous ORF. This is achieved by site-directed mutagenesis as follows:

(i) The termination codon TAA of HPV16 E6 in pIMS7 is altered using the oligonucleotide S20 to the sequence GGAA. This is in order to convert the normally separate ORFs for HPV16 E6 and E7 into a single ORF (pIMS7.1 - figure 2).

(ii) The termination codon TAA of HPV18 E6 in pIMS8 is altered using the oligonucleotide S21 to the sequence GGAA. This is in order to convert the normally separate ORFs for HPV18 E6 and E7 into a single ORF (pIMS8.1 - figure 2).

Abolition of the immortalising potential of E7

In order to destroy the immortalising properties of each of the E7 proteins, two key codons within the HPV16 E7 coding sequence, (cys24 and glu26 - figure 1(a)) and the equivalent codons from HPV18 E7 (cys27 and glu29 - figure 1(b)), are altered to glycine residues by site directed mutagenesis as follows.

(i) The sequence of the E7 gene is altered in pIMS7 to encode glycine at codons 24 and 26 (normally encoding cysteine and glutamate respectively, using oligonucleotide S22 (pIMS7.2 - figure 2).

(ii) The sequence of the E7 gene is altered in pIMS8 to encode glycine at codons 27 and 29 (normally encoding cysteine and glutamate respectively, using oligonucleotide S23 (pIMS8.1B - figure 2).

Reduction in intertypic recombination potential of HPV16 and HPV18 E6 and E7 sequence and elimination of potential vaccinia virus transcription termination signal

A potential difficulty with the presence of both HPV16 and HPV18 E6 and E7 specific DNA within the genome of a single virus, is that recombination between the two sets of related sequences could lead to loss or rearrangement of information such that expression of the required proteins is disrupted. The invention provides ways of minimising this risk. Firstly, by inserting the two sets of genetic information in the

vaccinia genome in opposite orientation to each other (so that recombination will result not in the loss of sequence information, but rather in its inversion). Secondly, by creating specific changes in the E6/7 sequence of one of the HPV virus strains at sites where the homology is greatest. These changes however are made in such a way that the amino acid coding potential of the genes remains unaltered.

The HPV18 E6 sequences is therefore altered by site-directed mutagenesis as follows:

The sequence TTTTTATTCTAGAATTAGAG (which begins 210 nucleotides from the start of E6 - underlined in figure 1(b)) is mutated, using oligonucleotide S24 to the sequence TTTCTACAGTAGAATCAGAG (pIMS8.2 - figure 2) (changed nucleotides are in bold type).

A second aim of this change is to eliminate from the HPV18 E6 sequence, the sequence TTTTTAT, which is a potential termination signal for the early vaccinia virus transcription enzyme (Rohrmann et al., Cell.,, 46, 1029, 1986).

Source and propagation of vaccinia virus

The Wyeth strain of vaccinia virus is used for construction of the therapeutic virus. It is propagated in Vero cells for the purposes of genetic manipulation, and in the human diploid fibroblast cell line MRC5 for the production of the final therapeutic virus stock.

Both cell lines are obtained from the National Institute of Biological Standards and Control, South Mims, U.K.. The Wyeth strain of vaccinia virus, vero cells and the cell line MRC5 are also available from the American Type Culture Collection, 12301 Parklawn Drive, Rockville, Maryland 20852, USA.

Identification and cloning of neutral sites from the vaccinia virus

Description of neutral sites

For the purpose of insertion of papillomavirus genes within the vaccinia virus genome, sites have been chosen to have two characteristics.

Firstly, they should be non-essential regions, i.e. insertion of foreign genes at these points, will not disrupt any functions of the vaccinia virus to the extent that the virus can no longer grow in tissue culture.

Secondly, they should be neutral sites, i.e. insertion of foreign genes all these points, will not increase or decrease the level of attenuation of the vaccinia virus.

The difference between these two factors can be seen by looking at the thymidine kinase (TK) gene of vaccinia virus. It is a non-essential region, and hence viruses with genes inserted in the TK gene can grow well in tissue culture (Mackett at al., J. Virol., 49, 857, 1984). However, such viruses have been found to be

greatly attenuated in vivo (Buller et al., Nature 317, 813, 1985). For the purpose of prophylactic vaccination, such increased attenuation might be desirable. However, for an immunotherapeutic strategy where the danger from the disease to be treated clearly outweighs the risk of vaccine associated complications, use of an attenuated virus is considered undesirable since it could compromise the immunological response to the papillomavirus antigens. Hence, the applicants have identified sites which they judge will not attenuate the virus any further, and have termed them 'neutral sites'. Such sites have been identified within the virus genome by careful analysis of the DNA sequence of the WR strain. The WR strain was originally derived from the Wyeth strain by passage in mouse brain. Therefore the two strains are closely related. The nucleotide sequence of three regions of the WR genome which contain the selected neutral sites are shown in figures 19,20 and 21. Four neutral sites (A-D) have been chosen on the above discussed criteria as follows:

Site A: gap between SalF17R and SalF19R

Site B: gap between SalF19R and SalF20.5R

Site C: in SalG2R, a potential non-functional gene

Site D: in HindB3.5R, a potential non-functional gene

These sites (A-D) can be identified by the following stretches of DNA sequence, each of which is 40 nucleotides in length.

A CTATCTACCAGATTATTATGTGTTATAAGGTACTTTTTCT

B TATTGTGCTACTGATTCTTCACAGACTGAAGATTGTTGAA

C TCTCTTAAAATGGTTGAGACCAAGCTTCGTTGTAGAAACA

D TGAGGCTACCTCGACATACGTGTGCGCTATCAAAGTGGAA

In other strains these sequences may vary, while still having substantial homology with those given above. In particular a site may have at least 90%, more preferably 95%, homology with the sequence given above.

Figures 3-5 show the distribution of initiation codons and open reading frames (ORFs) in the regions of the vaccinia virus genome shown in figures 19,20 and 21. Figures 6-8 show the same regions with a plot showing to what extent each reading frame conforms to the pattern of codon usage expected for vaccinia genes. A graph of codon usage is plotted for each of the three possible reading frames in each direction (Staden, R., Nucl. Acids Res., 12, 521, 1984; Staden, R., Nucl. Acids Res., 12, 551, 1984). In these codon usuage plots, the short vertical bar lines extending from the horizontal axes represent start codons. The longer vertical bar lines placed above the horizontal axes represent stop codons. This sort of plot is a useful way of helping to determine whether a particular ORF is a genuine vaccinia gene. Where there is a likely genuine gene, the graph of codon usuage will rise between a start codon and a stop codon. For example, in figure 7, it can be seen that the graph of codon usage rises over the region of the SalG2R ORF (the dotted line shows that this frame conforms most of the expected codon usage). For the

other two frames the graphs show that they do not conform to vaccinia codon usage. The peak labelled 'part of gk', and marked with a dashed line, also conforms well to vaccinia codon usage. In summary, a genuine gene must start with an initiation (start) codon, end with a termination (stop) codon, and should conform well to vaccinia codon usage along its length. In most cases the conformation to the vaccinia codon usage drops off sharply outside the gene.

The neutral sites are further described as follows

Site A. Gap between SalF17R and SalF19R

Site A is marked on figures 3 and 6. Figure 9 shows the actual DNA sequence with a translation of the ORFs on either side of the site. It can be seen that Site A is placed in an intergenic region between SalF17R and SalF19R. It is placed some 195 bases upstream of SalF19R to avoid any promoter elements associated with that gene. The sequence TTTTTCT (shown in italics) will act as a terminator of early RNA transcription for the SalF17R gene if it is an early gene. However, the site is placed downstream of the first of these, so it will not affect early termination of transcription if it occurs. Examination of figure 6 shows that there is no recognisable gene on the opposite strand at this point, and hence this sequence location is suitable as a neutral insertion site.

Site B. Gap between SalF19R and SalF20.5R

Site B is marked on figures 3 and 6. Figure 10 shows

the actual DNA sequence with a translation of the ORFs on either side of the site. It can be seen that Site B is placed in an intergenic region between SalF19R and SalF20.5R. Figure 6 shows that it is within a region of high vaccinia codon usage, but that this region does not form a genuine gene, having no initiation codon. In addition, figure 6 suggests that Sal20.5R is not a complete gene, as the conformation to vaccinia codon usage drops off dramatically at the start of the gene. In the event that SalF20.5R is a genuine gene, Site B is placed some 70 bases upstream of SalF20.5R which may well avoid any promoter elements associated with that gene. (Note: many vaccinia promoter elements are located in approximately 35 bases upstream of the start of the gene.) In addition SalF20R has no TTTTTNT (N=any nucleotide) transcription termination signal with which Site B could interfere. Hence this sequence location is suitable as a neutral insertion site.

Site C. Within SalG2R, a potential non-functional gene

Site C is marked on figure 4. The ORF SalG2R has considerable similarity to the guanylate kinase (GK) gene of yeast. This similarity is shown in figure 11. Sequence upstream of the SalG2R ORF (but in a different frame) has been added on to SalG2R, to see if the match to GK extends beyond the boundaries of the original open reading frame. The match appears to extend beyond the 5' end of the SalG2R ORF. In particular, an important

site in the yeast GK gene, the ATP/GTP binding site (shown underlined) only matches in the out of frame sequence upstream of the SalG2R ORF. Hence, it is very likely that the SalG2R gene is not active as a guanylate kinase and can be referred to as a 'pseudogene'. If the gene is inactive as the applicants deduce, then it will serve as a neutral insertion site.

Site D. Within HindB3.5R, a potential non-functional gene

Figure 5 shows that site D lies within the region designated HindB3.5R. This region, although conforming to vaccinia codon usage, has no start codon and is therefore not a genuine gene. The codon usage plot shown in figure 8 indicates that it probably was once a functional gene, and may well have been attached to HindB3R (a shift in the codon usage preference occurs here well away from the termination codon of the HindB3R ORF which suggests that the last section of HindB3R is not properly part of this gene.) Hence it is likely that HindB3.5R is not active as a gene and can be used as a neutral insertion site. Figure 12 shows the actual DNA sequence with a translation of the ORFs on either side of the site. It can be seen that site D is placed in an intergenic region between HindB3R and HindB4R as well as being within the non-functional HindB3.5R.

Preparation of vector for cloning of neutral sites

In order to insert foreign genetic information into the

neutral sites described above, DNA copies of the neutral sites, together with an appropriate amount of flanking DNA from the vaccinia genome (approximately 500 bases on either side) must first be cloned into a plasmid vector. These plasmids may then be used to introduce the foreign DNA into the vaccinia virus genome; the vaccinia virus 'flanking sequences' around the inserted gene serve to allow homologous recombination between the plasmid DNA and the viral DNA, with the consequent insertion of the foreign gene at the desired location.

Cloning of neutral site sequences

Plasmids containing flanking regions from the neutral sites are constructed as follows. DNA is prepared from the Wyeth strain of vaccinia virus by the method of Esposito et al., (J. Virol. Meth. 2: 175, 1981). The polymerase chain reaction (PCR) is used to remove an approximately 1000 base pair (bp) fragment from DNA of the Wyeth strain of vaccinia virus. Pairs of oligonucleotides are chosen approximately 500 bp either side of the chosen neutral site. These oligonucleotides are based on the sequence of the WR strain, but are chosen in regions where the sequence of the WR strain is identical to that of the Copenhagen strain (Goebel et al., Virology 179:247, 1990). The oligonucleotides incorporate restriction enzyme recognition sequences so that they can be cloned easily into a plasmid. For neutral sites A,B and D the restriction sites are EcoR1

and HindIII. For neutral site C the HindIII site is replaced by an Sphl site, since there is an internal HindIII site in the chosen flanking sequences.

The oligonucleotides used for PCR are listed below:

Site A leftMB 16

Site A rightMB 17

Site B leftMB 24

Site B rightMB 25

Site C leftMB 18

Site C rightMB 19

Site D leftMB 22

Site D rightMB 23

DNA fragments of approximately 1kb are then prepared using these pairs of oligonucleotides by PCR amplification, digested with EcoRI and HindIII (for site A,B and D) or with EcoRI and Sphl (for site C) and cloned into HindIII and EcoRI-digested pUC118 (figure 13) to generate the plasmids pIMMC7a, pIMMC7b, pIMMC7c and pIMMC7d.

Creation of unique restriction sites for insertion at the neutral sites

A suitable restriction enzyme site is then introduced at the selected location within each of the plasmids. This is achieved using site directed mutagenesis using an oligonucleotide containing the desired new unique site and flanked by 15 bases of sequence to either side

(see below). The plasmids modified in this fashion are designated pIMMC8a-d (figure 13).

original plasmid	oligonucleotidesite introduced	new plasmid
pIMMC7a	MB35SnaBl	pIMMC8a
pIMMC7b	MB36Hpal	pIMMCb
pIMMC7c	MB37Stul	pIMMC8c
pIMMC7d	MB38SnaBl	pIMMC8d

Cloning of the vaccinia virus early/late promoter sequences

The p7.5 and H6 promoters from vaccinia virus genomic DNA are prepared by PCR amplification as described below.

A pair of complementary oligonucleotides (S7 and S8) is synthesised to include the following restriction enzyme sites, HindIII, SnaBl, Hpal, HindIII, Sall, Ncol, Smal, SnaBl and EcoRl, such that the pair, after annealing, present at one end HindIII compatible overhanging ends, and at the other, EcoRl compatible overhanging ends. The two oligonucleotides are allowed to anneal and are inserted into pUC118 cut with EcoRl and HindIII (figure 14). The resulting vector is called pIMMC3.

A DNA molecule of approximately 180 bp containing the H6 promoter is removed from the WR strain of vaccinia virus by PCR amplification using the

oligonucleotides MB15 (anneals upstream and includes a 5'-SalI site) and MB7 (anneals downstream and includes a 5'-HindIII site). This is cloned into pIMMC3 cleaved with HindIII and SalI to create pIMMC4a (figure 14). A DNA molecule of approximately 200bp containing the p7.5 promoter is then removed from the WR strain of vaccinia virus by PCR amplification using the oligonucleotides MB32 (anneals upstream and includes a 5'-SalI site) and MB33 (anneals downstream and includes a 5'NcoI site). This is cloned into pIMMC3 cleaved with NcoI and SalI to create pIMMC14b.

Construction of the therapeutic virus

The strategy required to generate a recombinant vaccinia virus containing and expressing the E6-E7 proteins from HPV16 and HPV18, based on the elements described above involves five main stages as outlined below.

i) Cloning of the modified E6-7 genes downstream of vaccinia early promoter sequences

A DNA fragment containing the modified HPV16 E6-7 sequence is excised from pIMS7.2 by digestion with HindIII and SmaI, and cloned into HindIII and HpaI-digested pIMMC4a to generate pIMS12 (figure 15).

A DNA fragment containing the modified HPV18 E6-7 sequence is excised from pIMS8.2 by digestion with NcoI and SmaI, and cloned into NcoI and SmaI-digested pIMMC14b to generate pIMS14 (figure 15).

ii) Preparation of a plasmid vector containing both HPV16 and HPV18 E6-7 sequences together with their upstream vaccinia promoters.

A DNA fragment containing the HPV18 E6-7 region together with the upstream p7.5 promoter is excised from pIMS14 with SalI and SmaI and inserted into SalI and SmaI-digested pIMS12 to generate pIMS15 (figure 15)

iii) Insertion of the HPV E6-7/promoter "double" cartridge into the neutral site containing plasmids.

A DNA fragment containing both the HPV16 and HPV18 E6-7 coding regions together with their upstream promoter elements is excised from pIMS15 with SnaBl and inserted into the appropriately-digested neutral site-containing plasmids pIMMC7a-d. This step is shown in figure 16, and the resulting plasmids are designated pIMMC9a-d.

iv) Introduction via homologous recombination of the neutral site DNA, together with the intervening HPV sequences, into the vaccinia virus genome to create a recombinant virus expressing the two modified HPV E6-7 sequences.

The recombinant plasmids pIMMC9a-d are purified and allowed to recombine into vaccinia (figure 17) using standard protocols (Mackett et al., in D.M. Glover (ed) DNA Cloning: a Practical Approach, Oxford and Washington DC, IRL Press, 1985).

Viruses which have acquired the HPV sequences are identified by probing with radiolabelled HPV specific sequences. Viral plaques are lifted onto nitrocellulose (Villareal and Berg, Science 196, 183, 1977) and probed with radiolabelled NcoI-SmaI fragment from pIMS14 containing the HPV18 E67 gene. Recombinant viruses are then isolated from the agarose overlay and plaque purified three times. They are checked for the presence of the appropriate DNA sequences by Southern blotting of purified virus DNA using DNA probes derived from the HPV E6 and E7 genes, and for expression of the appropriate sequences by western blotting using antisera specific for the HPV E6 and E7 proteins.

Cloning of the therapeutic virus in MRC5 cells

Stocks of the final recombinant virus are prepared by growth in Vero cells, and are used to infect MRC5 cells deemed suitable for the preparation of material suitable for use as human vaccines. The virus is plaque-purified three times by standard methods, and finally a stock prepared for clinical use.

Confirmation of presence of the correct HPV DNA insert

A sample of this stock virus is checked once again for the presence of correctly configured virus DNA, and for expression of the correct virus

proteins. Figure 22 shows the analysis by PCR, of a recombinant vaccinia virus (v9a.1) in which the HPV DNA cassette is inserted at Site A. The diagram shown in panel (a) indicates the DNA fragments expected if insertion of the correct DNA has occurred. It can be seen in panel (b) that the actual pattern of PCR products generated is consistent with that expected.

Confirmation of expression of the HPV DNA insert

The recombinant viruses are then checked for expression of the expected HPV proteins. An example of this analysis is shown in Figure 23. Vero cells are infected with with recombinant virus v9a.1 (HPV DNA inserted at Site A), and the cells examined by western blotting for the presence of the HPV E67 fusions proteins using monoclonal antibodies specific for the HPV16 E7 protein (camvir3) and for the HPV18 E7 protein (7E10). It can be seen that both monoclonal antibodies recognise specifically proteins of the expected size in cells infected with the recombinant virus v9a.1, but not in cells infected with the control parent virus Wyeth strain. These recognised proteins co-migrate with proteins synthesised by in vitro translation of mRNA encoding the expected HPV fusion proteins (HPV 16 E67 and HPV18 E67). This experiment indicates

successful expression of the heterologous gene sequences from the recombinant virus.

Stability of the HPV DNA insert

For the recombinant virus to be of use clinically, it is important that the inserted sequences remain genetically stable over multiple virus passage, and the DNA insert was carefully designed to promote this genetic instability. To confirm the stability of the HPV information within the recombinant virus genome, the virus is subjected to 9 serial passages, (multiplicity of infection =10pfu/cell) in Vero cells. Subsequently 20 plaque isolates are picked, and analysed for the presence of the correct HPV DNA insert by PCR analysis as described in Figure 22. The data obtained for recombinant virus A are shown in Figure 24. All 20 virus isolates retained the HPV information in the expected genetic arrangement indicating a considerable degree of genetic stability.

Animal experiments

The virulence of the recombinant virus is compared in animal experiments with that of the parental Wyeth strain. Groups of 20 mice are inoculated intranasally each with 10^7 pfu of Wyeth strain or recombinant virus in a total volume of 20µl. Two mice are sacrificed at 1 day, 3 days and 5days

following inoculation, and the lungs dissected out. The amount of virus present in the lungs is then measured by grinding the tissue, and assay of the homogenate by standard vaccinia virus plaque assay. The results of such an experiment for the recombinant virus v9a.1 (HPV infromation inserted at site A) are shown in Figure 25. It can be seen that the recombinant virus retains the ability to replicate in mice, and that the level of virus produced in the lungs of the infected animals is similar to that seen with the parental Wyeth strain.

Therapeutic Use

A stock of the recombinant virus is prepared by infection of MRC5 cells, and adjusted to a concentration of not less than 10^8 pfu/ml. 20 µl of this material is applied to the arm of the patient, which is then scarified through the virus droplet with a bifurcated needle, according to the standard procedure used for vaccination against smallpox.

CLAIMS

1. A recombinant virus vector for use as an immunotherapeutic or vaccine which comprises at least one pair of nucleotide sequences heterologous to said virus which have sufficient sequence homology that recombination between them might be expected;

wherein said pair of nucleotide sequences are arranged in said virus vector such that they are inverted with respect to each other to reduce the likelihood of recombination events leading to loss of part or all of said sequences and said virus vector is able to infect a mammalian host cell and express as polypeptide the heterologous nucleotide sequences in said host cell.

2. A recombinant virus vector according to claim 1 wherein at least one pair of nucleotide sequences encode part or all of human papillomavirus (HPV) wild-type proteins or mutant proteins immunologically cross-reactive with said wild-type proteins.

3. A recombinant virus vector according to claim 1 or claim 2 wherein the pair of nucleotide sequences encode part or all of the HPV wild-type proteins HPV16E7 and HPV18E7 or mutant proteins immunologically cross-reactive therewith.

4. A recombinant virus vector according to claim 1 or claim 2 wherein the pair of nucleotide sequences encode part or all of the HPV wild-type proteins HPV16E6 and HPV18E6 or mutant proteins immunologically cross-reactive therewith.

5. A recombinant virus vector according to claim 3 which comprises a further pair of nucleotide sequences which encode part or all of the HPV wild-type proteins HPV16E6 and HPV18E6 or mutant proteins immunologically cross-reactive therewith.

6. A recombinant virus vector according to any one of the preceding claims wherein two or more nucleotide sequences of different said pairs may be fused together to form a single open reading frame.

7. A recombinant virus vector according to claim 6 wherein the fusions are via a single codon encoding a small neutral amino acid.

8. A recombinant virus vector according to claim 7 wherein the amino acid is glycine.

9. A recombinant virus vector according to claim 5 in which the pairs of nucleotide sequences are arranged in the virus vector according to any one of the options as shown in Figure 26.

10. A recombinant virus vector according to claim 9 which comprises

a first open reading frame having a fused genetic sequence encoding part or all of the wild-type proteins E6 and E7 from HPV16; and

a separate second open reading frame having a fused genetic sequence encoding part or all of the wild-type proteins E6 and E7 from HPV18;

wherein the first and second open reading frames may be inverted with respect to one another whereby either:

i) the E6 coding sequences of HPV16 and HPV18 are both located between the E7 coding sequences of HPV16 and HPV18; or

ii) the E7 coding sequences of HPV16 and HPV18 are both located between the E6 coding sequences of HPV16 and HPV18; and

wherein any of said wild-type proteins may be replaced by a mutant protein immunologically cross-reactive therewith.

11. A recombinant virus vector according to claim 10 wherein each of the first and second open reading frames has a corresponding promoter and the two open reading frames each with its promoter, are arranged next to each other in the virus.

12. A recombinant virus vector according to claim 11 wherein either:

i) the promoters are located between the first and second open reading frames whereby the open reading frames are transcribed outwardly; or

ii) the promoters are located outside the first and second open reading frames whereby the open reading frames are transcribed inwardly.

13. A recombinant virus vector according to any one of claims 9 to 12 which comprises

a first open reading frame having a fused genetic sequence encoding part or all of the wild-type proteins E6 and E7 from HPV16; and

a separate second open reading frame having a fused genetic sequence encoding part or all of the wild-type proteins E6 and E7 from HPV18;

wherein the E6 coding sequences of HPV16 and HPV18 are both located between the E7 coding sequences of HPV16 and HPV18; and

each open reading frame has a corresponding promoter, the promoters being located between the first and second open reading frames whereby the open reading frames are transcribed outwardly; and

wherein any of said wild-type proteins may be replaced by a mutant protein immunologically cross-reactive therewith.

14. A recombinant virus vector according to any one of the preceding claims wherein either or both of the nucleotide sequences in a said pair of nucleotide

sequences are altered to make them less homologous than an equivalent pair of nucleotide sequences encoding wild-type HPV proteins.

15. A recombinant virus vector according to claim 14 wherein the alteration in nucleotide sequence does not result in an alteration of the encoded amino acid sequence.

16. A recombinant virus vector according to claim 3 wherein the wild-type proteins HPV16E7 and HPV18E7 are replaced with mutant proteins which are substantially homologous to said wild-type proteins and in which the residues cys 24 and glu 26 of wild-type protein HPV16E7 and the residues cys 27 and glu 29 of wild-type protein HPV18E7 are replaced with glycine residues.

17. A recombinant virus vector according to any one of the preceding claims wherein said heterologous nucleotide sequences may comprise part or all of the sequences shown in Figures 1(a) and 1(b).

18. A recombinant virus vector according to any one of the preceding claims which is derivable from vaccinia virus.

19. A recombinant virus vector according to any one of the preceding claims wherein the nucleotide

sequences are inserted into the virus vector at one or more neutral sites, the disruption of which by the insertion of the nucleotide sequences does not substantially adversely affect viral functions relating to the replicative ability of the virus in the mammalian cell.

20. A recombinant virus vector according to claim 19 which is derivable from vaccinia virus and wherein the neutral sites may be one or more of:

A) the gap between SalIF17R and SalIF19R of strain WR comprising at least part of the sequence CTATCTACCAGATTATTATGTGTTATAAGGTACTTTTTCT;

B) the gap between SalIF19R and SalIF20.5R of strain WR comprising at least part of the sequence TATTGTGCTACTGATTCTTCACAGACTGAAGATTGTTGAA;

C) a region in SalIG2R of strain WR comprising at least part of the sequence TCTCTTAAAATGGTTGAGACCAAGCTTCGTTGTAGAAACA;

D) a region in HindB3.5R of strain WR comprising at least part of the sequence TGAGGCTACCTCGACATACGTGTGCGCTATCAAAGTGGAA;

E) a sequence having at least 90% sequence homology to those sequences A) to D) identified above.

21. A method for making a recombinant virus vector according to claim 19 or claim 20 which comprises inserting a said heterologous nucleotide sequences into one or more neutral sites in a virus vector, the

disruption of such a site by said insertion will not substantially adversely affect the replicative ability of the virus and wherein the neutral site has been previously identified by: (a) analysing a viral genome to identify open reading frames which are likely to encode functional genes; and (b) selecting sites between open reading frames for functional genes or sites within sequences for non-functional genes.

22. A recombinant virus vector obtainable by the method of claim 21.

23. A method which comprises using a recombinant virus vector according to any one of claims 1 to 20 or to claim 22 to manufacture a medicament for use as an immunotherapeutic or vaccine against a condition thought to be caused by HPV infection, for example cervical cancer.

24. A method which comprises using a recombinant virus vector according to any one of claims 1 to 20 or to claim 22 to specifically activate cells of the immune system to HPV proteins.

Fig. 1a.

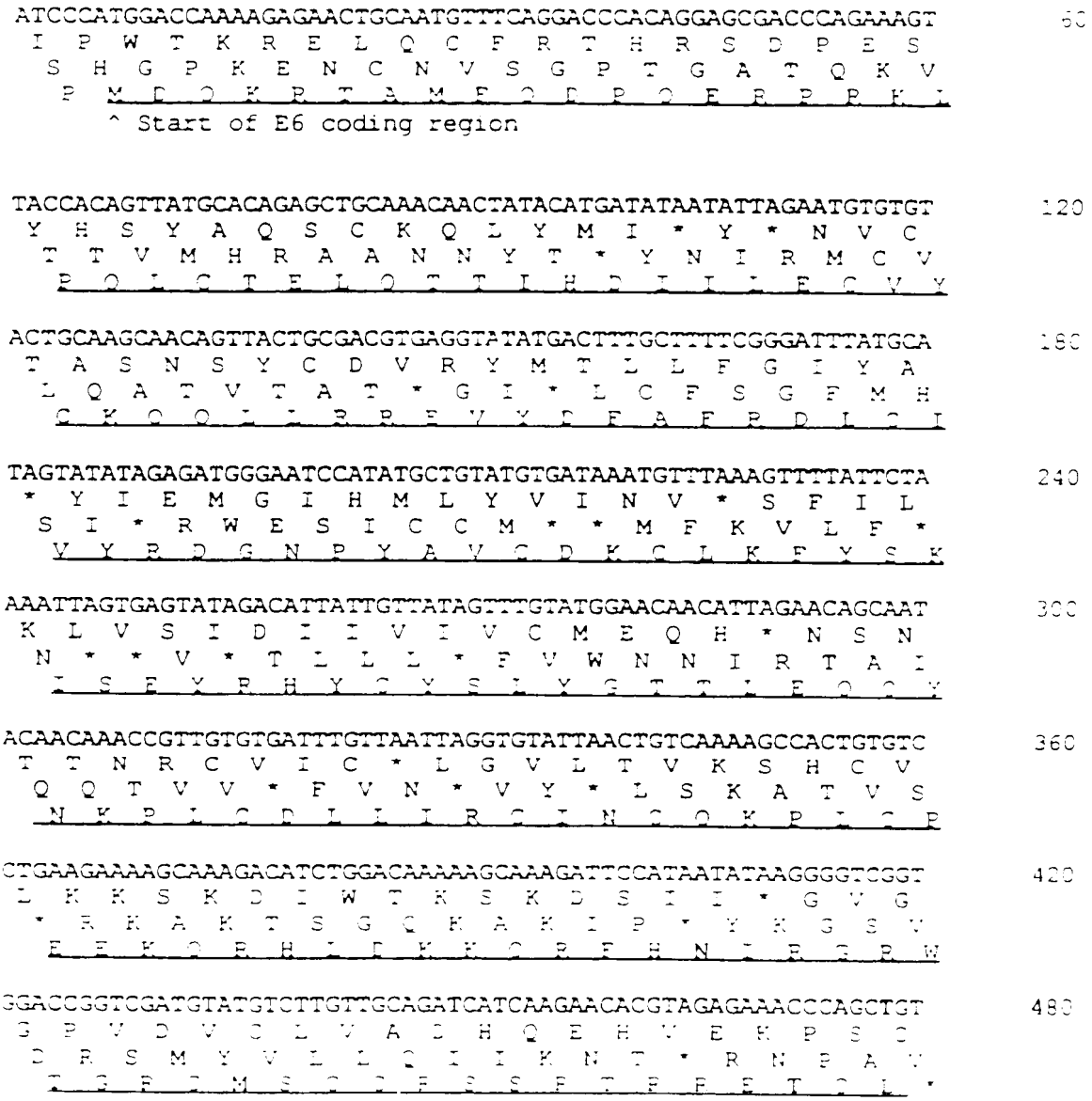

ATCCCATGGACCAAAAGAGAACTGCAATGTTTCAGGACCCACAGGAGCGACCCAGAAAGT 60
I P W T K R E L Q C F R T H R S D P E S
S H G P K E N C N V S G P T G A T Q K V
P M D Q K R T A M F Q D P Q E R P R K L
^ Start of E6 coding region
TACCACAGTTATGCACAGAGCTGCAAACAACTATACATGATATAATATTAGAATGTGTGT 120
Y H S Y A Q S C K Q L Y M I * Y * N V C
T T V M H R A A N N Y T * Y N I R M C V
P Q L C T E L Q T T I H D I I L E C V Y
ACTGCAAGCAACAGTTACTGCGACGTGAGGTATATGACTTTGCTTTTCGGGATTTATGCA 180
T A S N S Y C D V R Y M T L L F G I Y A
L Q A T V T A T * G I * L C F S G F M H
C K Q Q L L R R E V Y D F A F R D L C I
TAGTATATAGAGATGGGAATCCATATGCTGTATGTGATAAATGTTTAAAGTTTTATTCTA 240
* Y I E M G I H M L Y V I N V * S F I L
S I * R W E S I C C M * * M F K V L F *
V Y R D G N P Y A V C D K C L K F Y S K
AAATTAGTGAGTATAGACATTATTGTTATAGTTTGTATGGAACAACATTAGAACAGCAAT 300
K L V S I D I I V I V C M E Q H * N S N
N * * V * T L L L * F V W N N I R T A I
I S E Y R H Y C Y S L Y G T T L E Q Q Y
ACAACAAACCGTTGTGTGATTTGTTAATTAGGTGTATTAACTGTCAAAAGCCACTGTGTC 360
T T N R C V I C * L G V L T V K S H C V
Q Q T V V * F V N * V Y * L S K A T V S
N K P L C D L I I R C I N C Q K P L C P
CTGAAGAAAAGCAAAGACATCTGGACAAAAAGCAAAGATTCCATAATATAAGGGGTCGGT 420
L K K S K D I W T K S K D S I I * G V G
* R K A K T S G Q K A K I P * Y K G S V
E E K Q R H L D K K Q R F H N I R G R W
GGACCGGTCGATGTATGTCTTGTTGCAGATCATCAAGAACACGTAGAGAAACCCAGCTGT 480
G P V D V C L V A D H Q E H V E K P S C
D R S M Y V L L Q I I K N T * R N P A V
T G R I M S C I F S S F T R E E T Q L *

Fig. 1a (Cont).

```
AATCATGCATGGAGATACACCTACATTGCATGAATATATGTTAGATTTGCAACCAGAGAC      540
 N  H  A  W  R  Y  T  Y  I  A  *  I  Y  V  R  F  A  T  R  D
 I  M  H  G  D  T  P  T  L  H  E  Y  M  L  D  L  Q  P  E  T
 S  C  M  E  I  H  L  H  C  M  N  I  C  *  I  C  N  Q  R  Q
    ^ Start of E7 coding region

AACTGATCTCTACTGTTATGAGCAATTAAATGACAGCTCAGAGGAGGAGGATGAAATAGA      600
 N  *  S  L  L  L  *  A  I  K  *  Q  L  R  G  G  G  *  N  R
 T  D  L  Y  C  Y  E  Q  L  N  D  S  S  E  E  E  D  E  I  D
    L  I  S  T  V  M  S  N  *  M  T  A  Q  R  R  R  M  K  *  M

TGGTCCAGCTGGACAAGCAGAACCGGACAGAGCCCATTACAATATTGTAACCTTTTGTTG      660
 W  S  S  W  T  S  R  T  G  Q  S  P  L  Q  Y  C  N  L  L  L
 G  P  A  G  Q  A  E  P  D  R  A  H  Y  N  I  V  T  F  C  C
    V  Q  L  D  K  Q  N  R  T  E  P  I  T  I  L  *  P  F  V  A

CAAGTGTGACTCTACGCTTCGGTTGTGCGTACAAAGCACACACGTAGACATTCGTACTTT      720
 Q  V  *  L  Y  A  S  V  V  R  T  K  H  T  R  R  H  S  Y  F
 K  C  D  S  T  L  R  L  C  V  Q  S  T  H  V  D  I  R  T  L
    S  V  T  L  R  F  G  C  A  Y  K  A  H  T  *  T  F  V  L  W

GGAAGACCTGTTAATGGGCACACTAGGAATTGTGTGCCCCATCTGTTCTCAGAAACCATA      780
 G  R  P  V  N  G  H  T  R  N  C  V  P  H  L  F  S  E  T  I
 E  D  L  L  M  G  T  L  G  I  V  C  P  I  C  S  Q  K  P  *
    K  T  C  *  W  A  H  *  E  L  C  A  P  S  V  L  R  N  H  N

ACCCGGGTGA                                                        840
 T  R  V
 P  G  *
 P  G
```

Fig.1b.

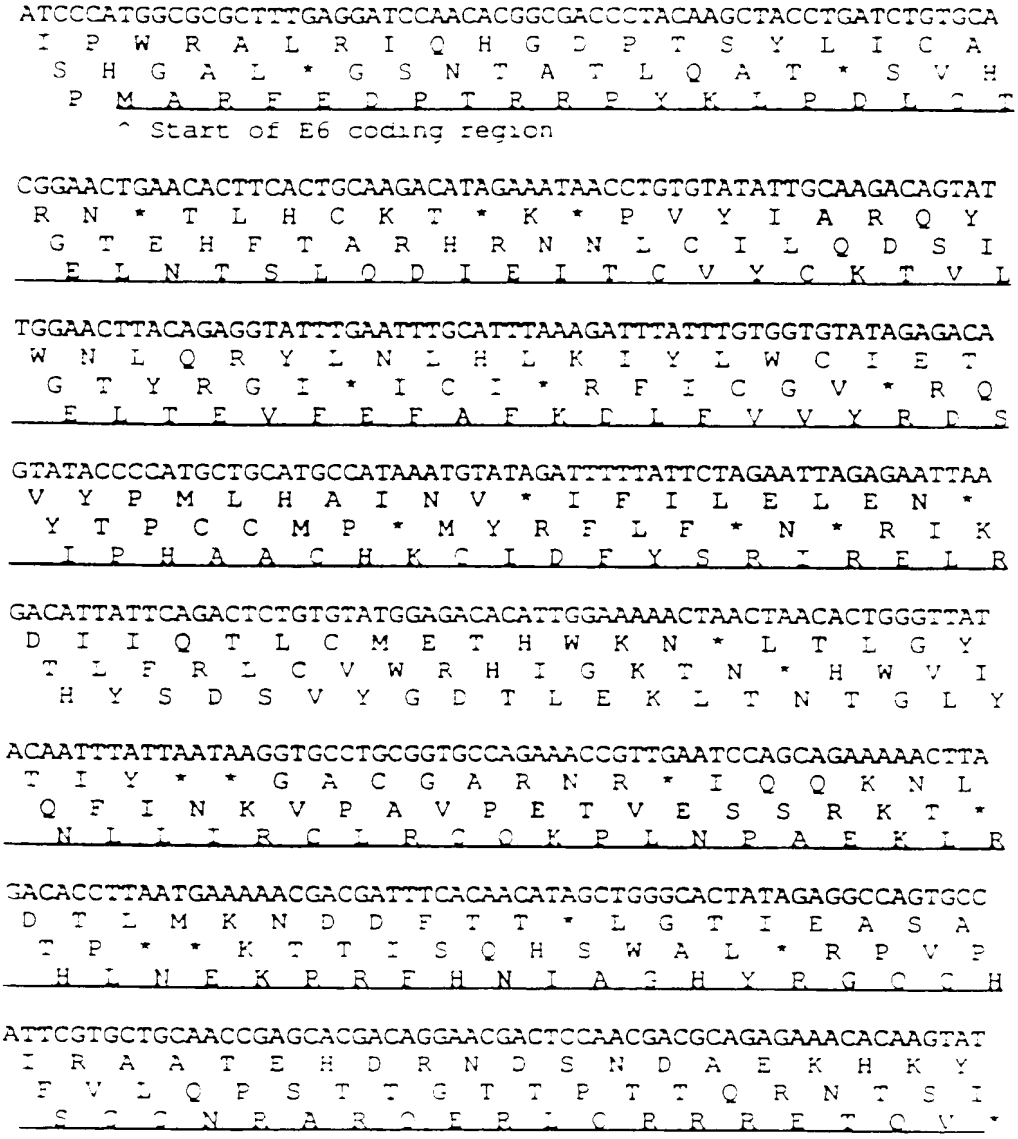

ATCCCATGGCGCGCTTTGAGGATCCAACACGGCGACCCTACAAGCTACCTGATCTGTGCA 60
I P W R A L R I Q H G D P T S Y L I C A
S H G A L * G S N T A T L Q A T * S V H
P M A R F E D P T R R P Y K L P D L C T
^ Start of E6 coding region
CGGAACTGAACACTTCACTGCAAGACATAGAAATAACCTGTGTATATTGCAAGACAGTAT 120
R N * T L H C K T * K * P V Y I A R Q Y
G T E H F T A R H R N N L C I L Q D S I
E L N T S L Q D I E I T C V Y C K T V L
TGGAACTTACAGAGGTATTTGAATTTGCATTTAAAGATTTATTTGTGGTGTATAGAGACA 180
W N L Q R Y L N L H L K I Y L W C I E T
G T Y R G I * I C I * R F I C G V * R Q
E L T E V F E F A F K D L F V V Y R D S
GTATACCCCATGCTGCATGCCATAAATGTATAGATTTTTATTCTAGAATTAGAGAATTAA 240
V Y P M L H A I N V * I F I L E L E N *
Y T P C C M P * M Y R F L F * N * R I K
I P H A A C H K C I D F Y S R I R E L R
GACATTATTCAGACTCTGTGTATGGAGACACATTGGAAAAACTAACTAACACTGGGTTAT 300
D I I Q T L C M E T H W K N * L T L G Y
T L F R L C V W R H I G K T N * H W V I
H Y S D S V Y G D T L E K L T N T G L Y
ACAATTTATTAATAAGGTGCCTGCGGTGCCAGAAACCGTTGAATCCAGCAGAAAAACTTA 360
T I Y * * G A C G A R N R * I Q Q K N L
Q F I N K V P A V P E T V E S S R K T *
N L L I R C L R C Q K P L N P A E K L R
GACACCTTAATGAAAAACGACGATTTCACAACATAGCTGGGCACTATAGAGGCCAGTGCC 420
D T L M K N D D F T T * L G T I E A S A
T P * * K T T I S Q H S W A L * R P V P
H L N E K P R F H N I A G H Y R G C C H
ATTCGTGCTGCAACCGAGCACGACAGGAACGACTCCAACGACGCAGAGAAACACAAGTAT 480
I R A A T E H D R N D S N D A E K H K Y
F V L Q P S T T G T T P T T Q R N T S I
S C C N R A R Q E R L C P R R R E T Q V *

Fig.1b(Cont).

```
AATATTAAGTATGCATGGACCTAAGGCAACATTGCAAGACATTGTATTGCATTTAGAGCC      540
 N  I  K  Y  A  W  T  *  G  N  I  A  R  H  C  I  A  F  R  A
  I  L  S  M  H  G  P  K  A  T  L  Q  D  I  V  L  H  L  E  P
   Y  *  V  C  M  D  L  R  Q  H  C  K  T  L  Y  C  I  *  S  P
          ^ Start of E7 coding region

CCAAAATGAAATTCCGGTTGACCTTCTATGTCACGAGCAATTAAGCGACTCAGAGGAAGA      600
 P  K  *  N  S  G  *  P  S  M  S  R  A  I  K  R  L  R  G  R
  Q  N  E  I  P  V  D  L  L  C  H  E  Q  L  S  D  S  E  E  E
   K  M  K  F  R  L  T  F  Y  V  T  S  N  *  A  T  Q  R  K  K

AAACGATGAAATAGATGGAGTTAATCATCAACATTTACCAGCCCGACGAGCCGAACCACA      660
 K  R  *  N  R  W  S  *  S  S  T  F  T  S  P  T  S  R  T  T
  N  D  E  I  D  G  V  N  H  Q  H  L  P  A  R  R  A  E  P  Q
   T  M  K  *  M  E  L  I  I  N  I  Y  Q  P  D  E  P  N  H  N

ACGTCACACAATGTTGTGTATGTGTTGTAAGTGTGAAGCCAGAATTGAGCTAGTAGTAGA      720
 T  S  H  N  V  V  Y  V  L  *  V  *  S  Q  N  *  A  S  S  R
  R  H  T  M  L  C  M  C  C  K  C  E  A  R  I  E  L  V  V  E
   V  T  Q  C  C  V  C  V  V  S  V  K  P  E  L  S  *  *  *  K

AAGCTCAGCAGACGACCTTCGAGCATTCCAGCAGCTGTTTCTGAACACCCTGTCCTTTGT      780
 K  L  S  R  R  P  S  S  I  P  A  A  V  S  E  H  P  V  L  C
  S  S  A  D  D  L  R  A  F  Q  Q  L  F  L  N  T  L  S  F  V
   A  Q  Q  T  T  F  E  H  S  S  S  C  F  *  T  P  C  P  L  C

GTGTCCGTGGTGTGCATCCCAGCAGTAACCCGGGTGA                              840
 V  S  V  V  C  I  P  A  V  T  R  V
  C  P  W  C  A  S  Q  Q  *  P  G  *
   V  R  G  V  H  P  S  S  N  P  G
```

Fig. 2.

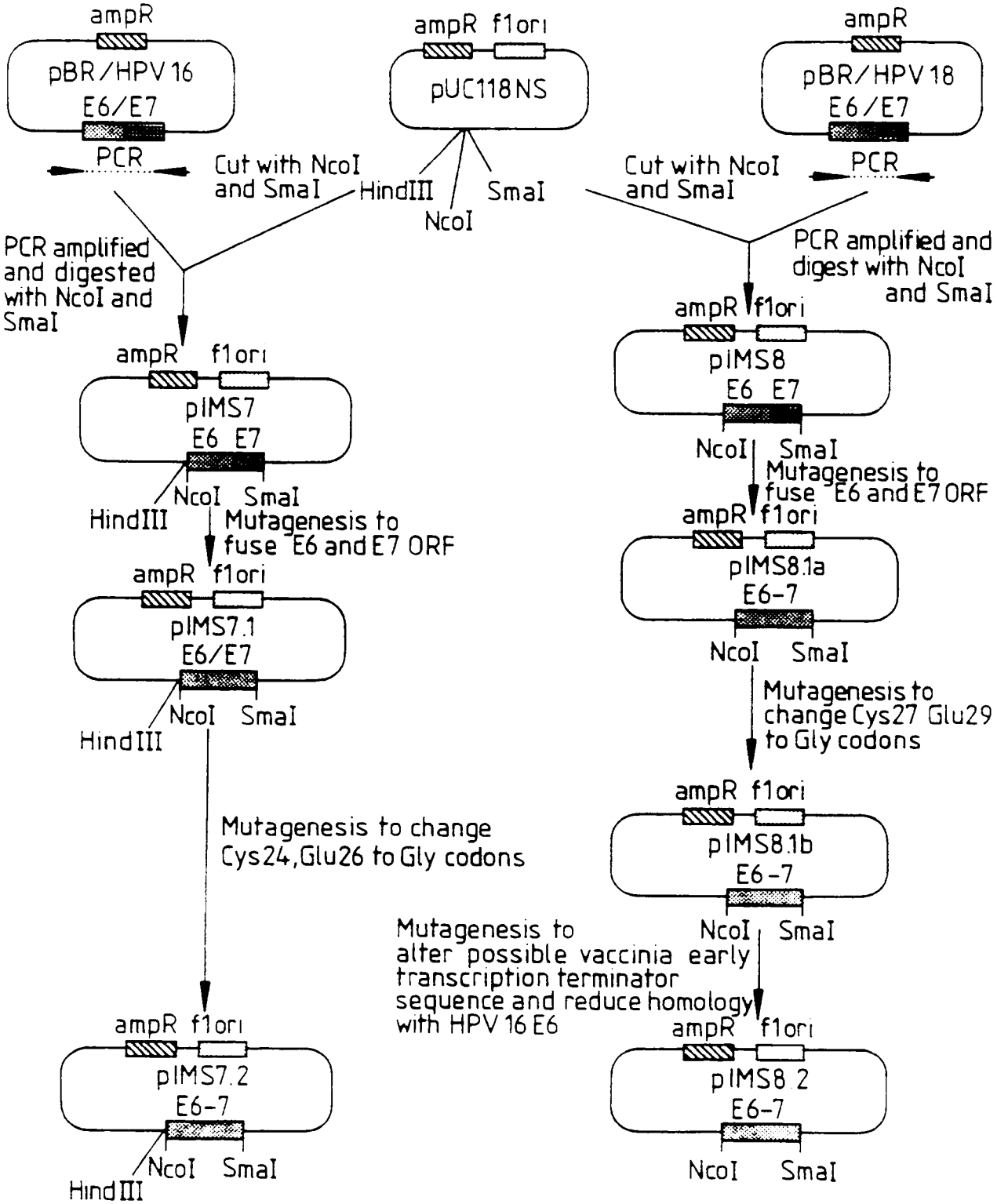

ampR
pBR/HPV16
E6/E7
PCR
Cut with NcoI
and SmaI
ampR f1ori
pUC118NS
HindIII
SmaI
NcoI
ampR
pBR/HPV18
E6/E7
Cut with NcoI
and SmaI
PCR
PCR amplified
and digested
with NcoI and
SmaI
PCR amplified and
digest with NcoI
and SmaI
ampR f1ori
pIMS7
E6 E7
HindIII
NcoI SmaI
Mutagenesis to
fuse E6 and E7 ORF
ampR f1ori
pIMS7.1
E6/E7
HindIII
NcoI SmaI
Mutagenesis to change
Cys24, Glu26 to Gly codons
ampR f1ori
pIMS7.2
E6-7
HindIII
NcoI SmaI
ampR f1ori
pIMS8
E6 E7
NcoI SmaI
Mutagenesis to
fuse E6 and E7 ORF
ampR f1ori
pIMS8.1a
E6-7
NcoI SmaI
Mutagenesis to
change Cys27 Glu29
to Gly codons
ampR f1ori
pIMS8.1b
E6-7
NcoI SmaI
Mutagenesis to
alter possible vaccinia early
transcription terminator
sequence and reduce homology
with HPV16 E6
ampR f1ori
pIMS8.2
E6-7
NcoI SmaI

Fig. 3.

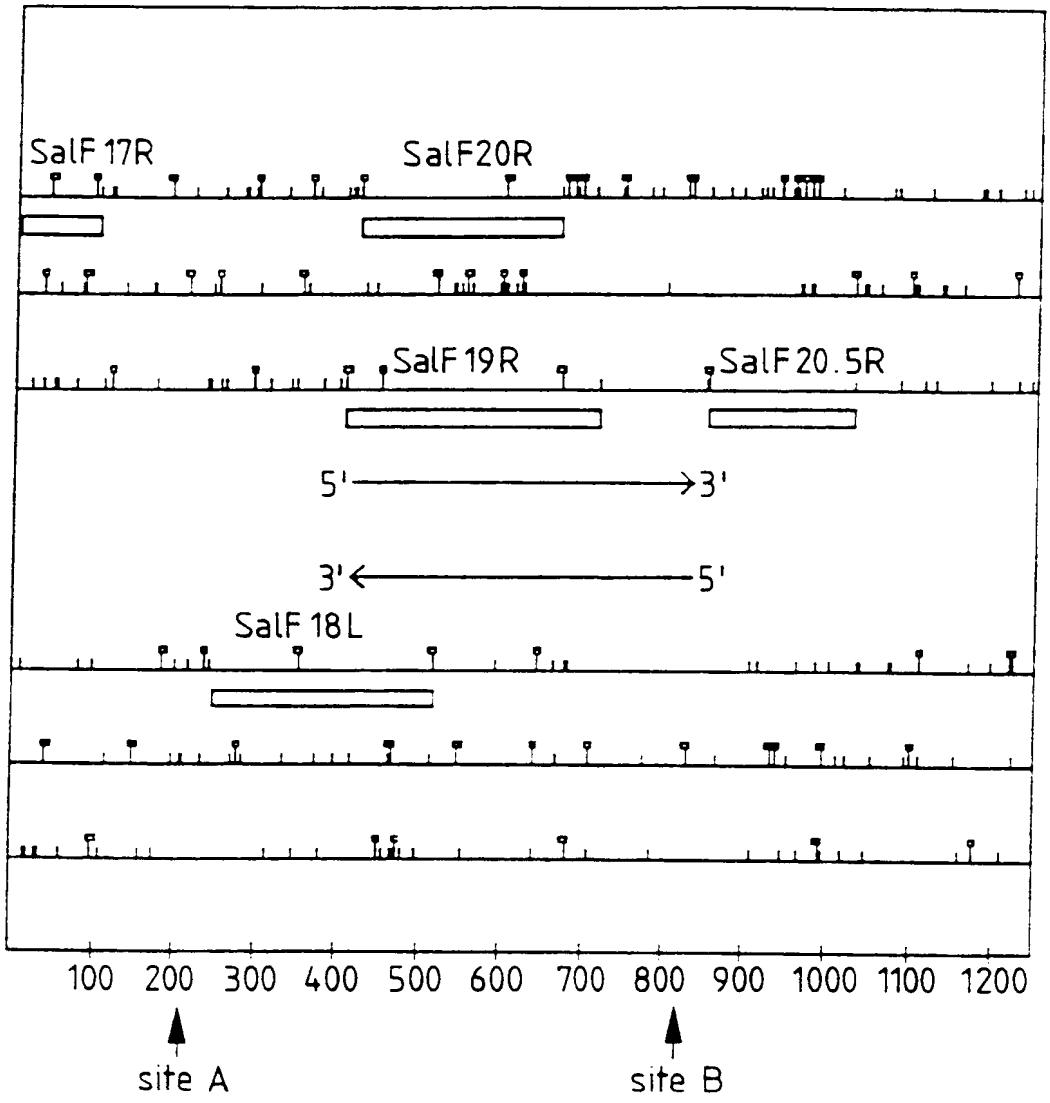

SalF 17R
SalF20R
SalF 19R
SalF 20.5R
5'
3'
3'
5'
SalF 18L
100 200 300 400 500 600 700 800 900 1000 1100 1200
site A
site B

Fig.4.

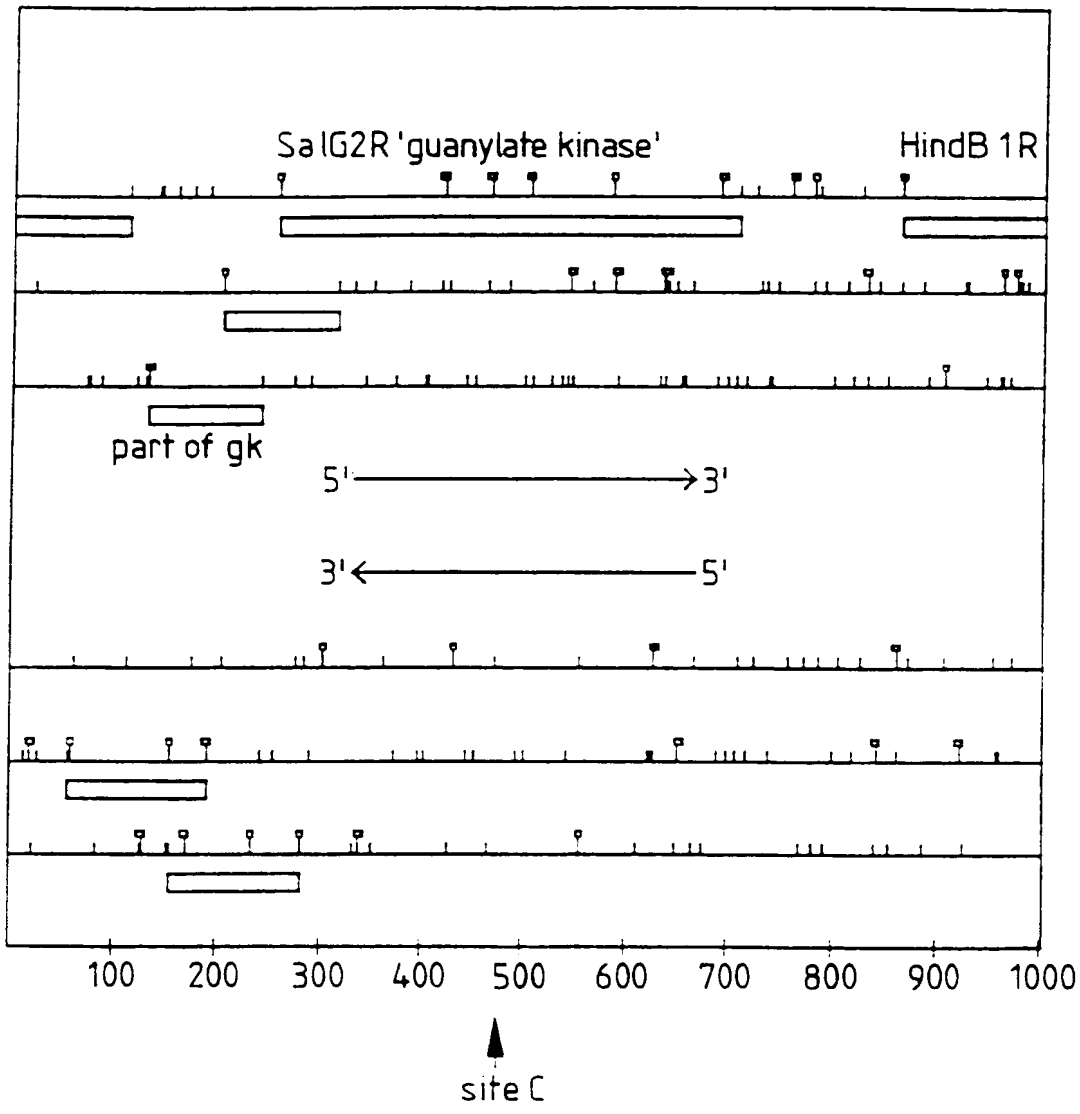

SalG2R 'guanylate kinase'
HindB 1R
part of gk
5'
3'
3'
5'
100
200
300
400
500
600
700
800
900
1000
site C

Fig. 5.

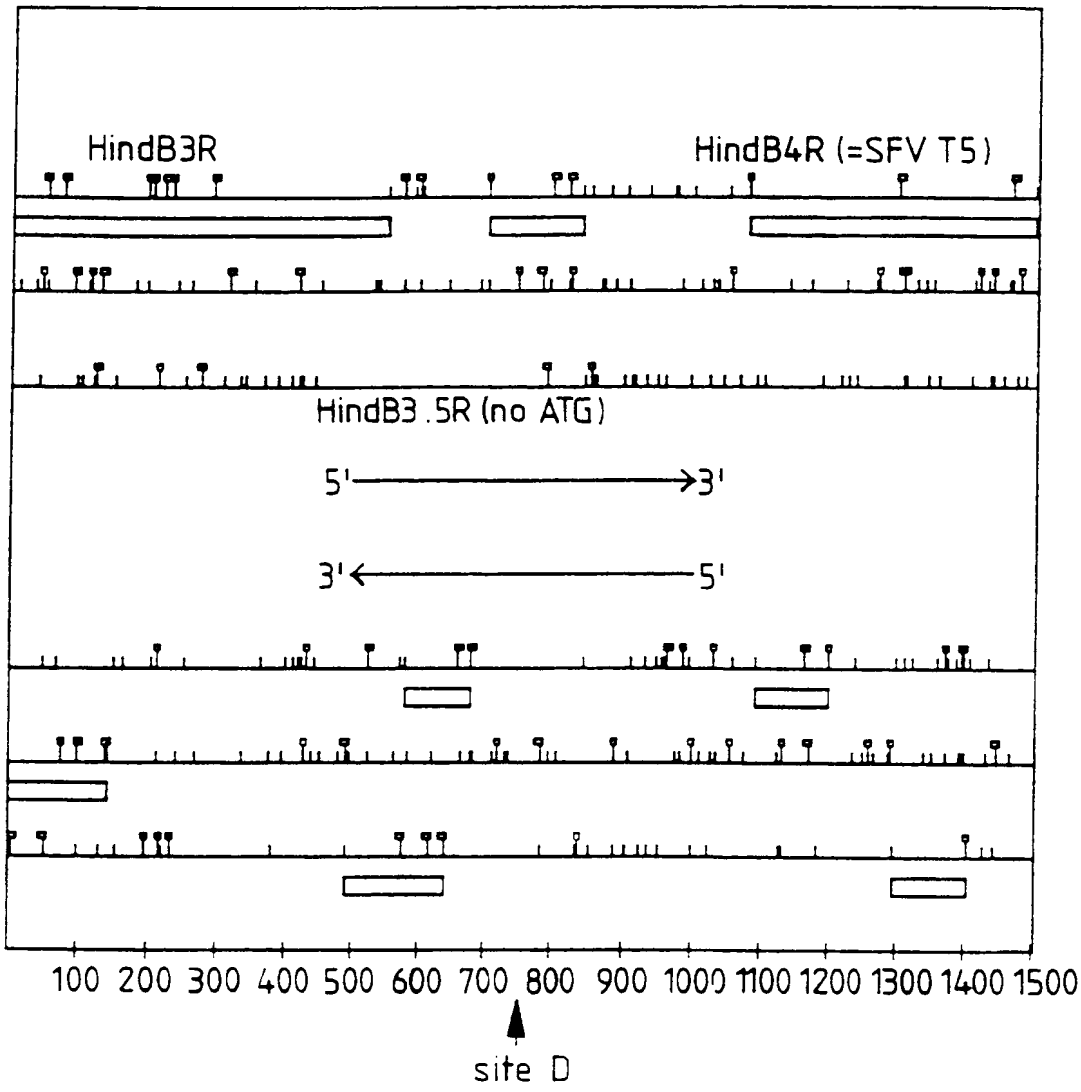

HindB3R
HindB4R (=SFV T5)
HindB3 .5R (no ATG)
5'
3'
3'
5'
100 200 300 400 500 600 700 800 900 1000 1100 1200 1300 1400 1500
site D

Fig. 6.

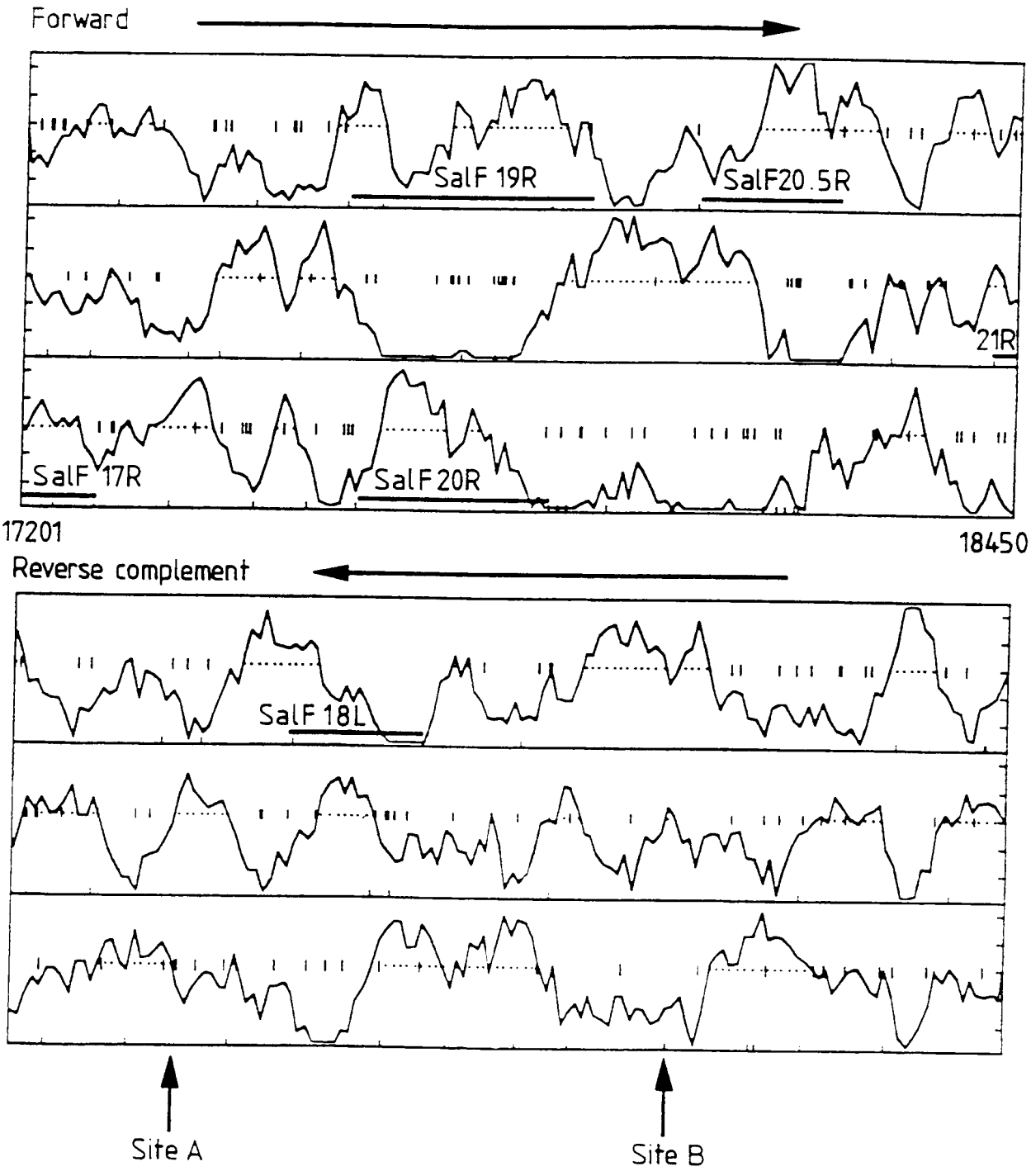

Forward
SalF 19R
SalF20.5R
21R
SalF 17R
SalF 20R
17201
18450
Reverse complement
SalF 18L
Site A
Site B

Fig. 7.

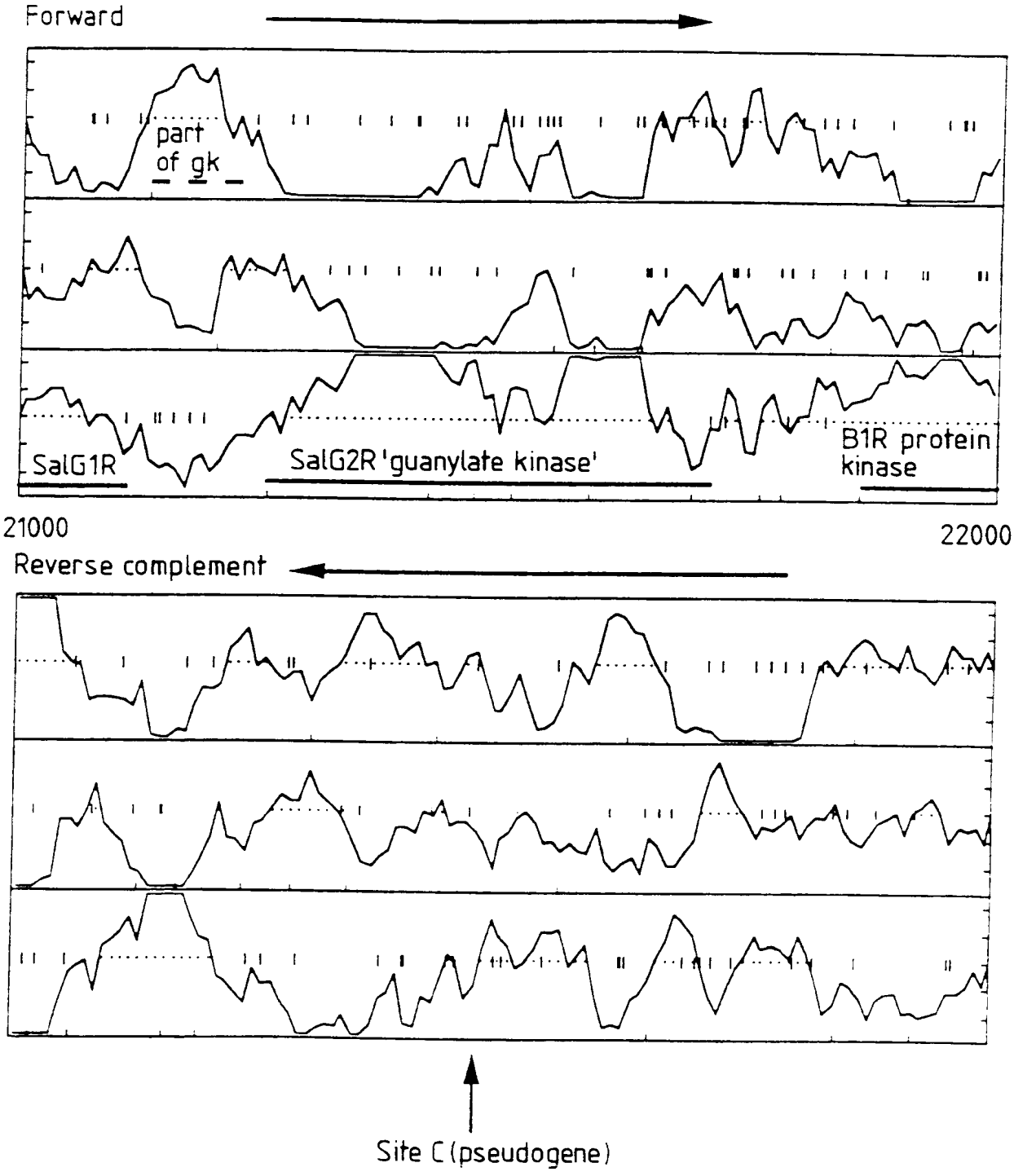

Forward
part
of gk
SalG1R
SalG2R 'guanylate kinase'
B1R protein
kinase
21000
22000
Reverse complement
Site C (pseudogene)

Fig. 8.

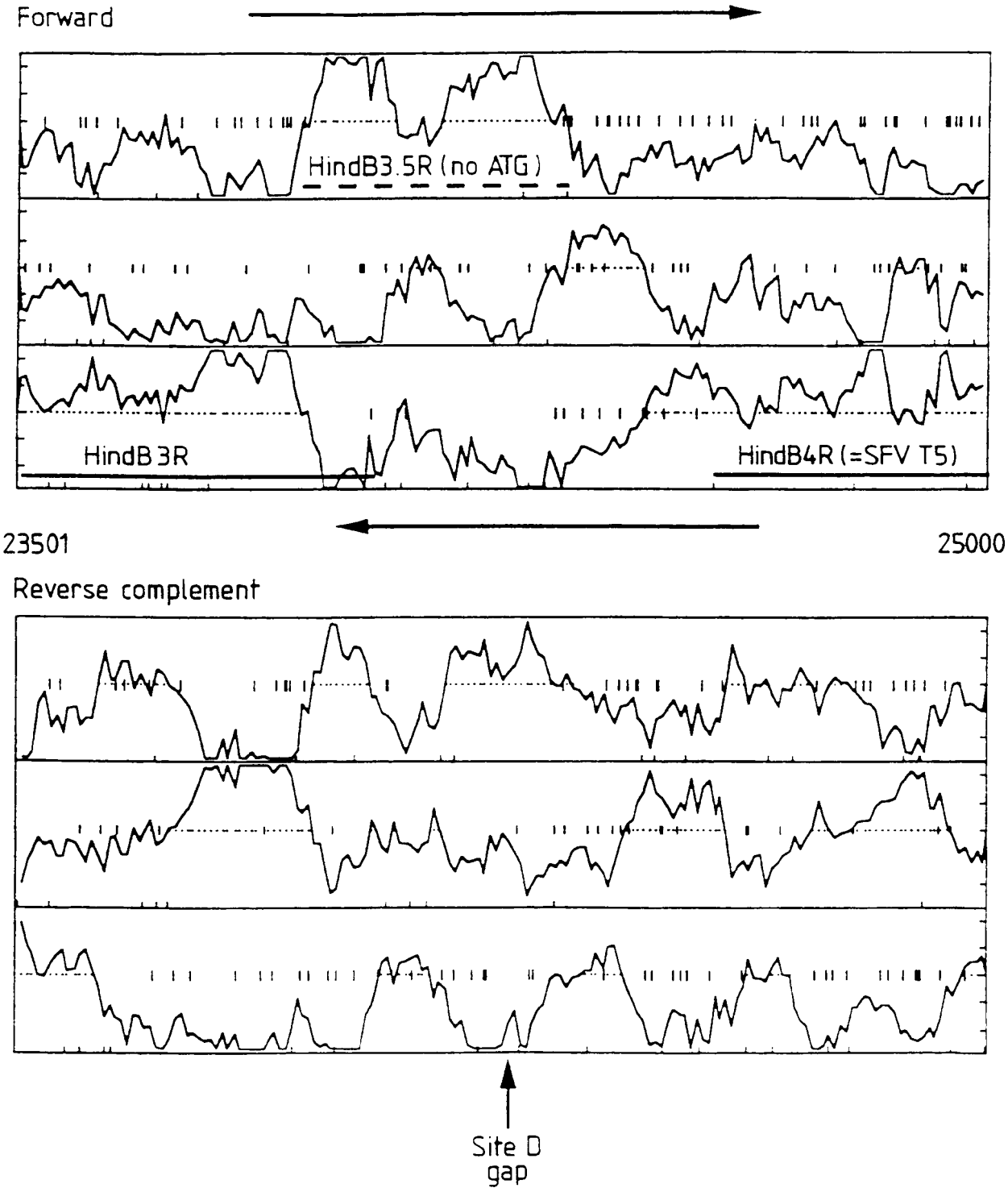

Forward
HindB3.5R (no ATG)
HindB3R
HindB4R (=SFV T5)
23501
25000
Reverse complement
Site D
gap

Fig. 9.

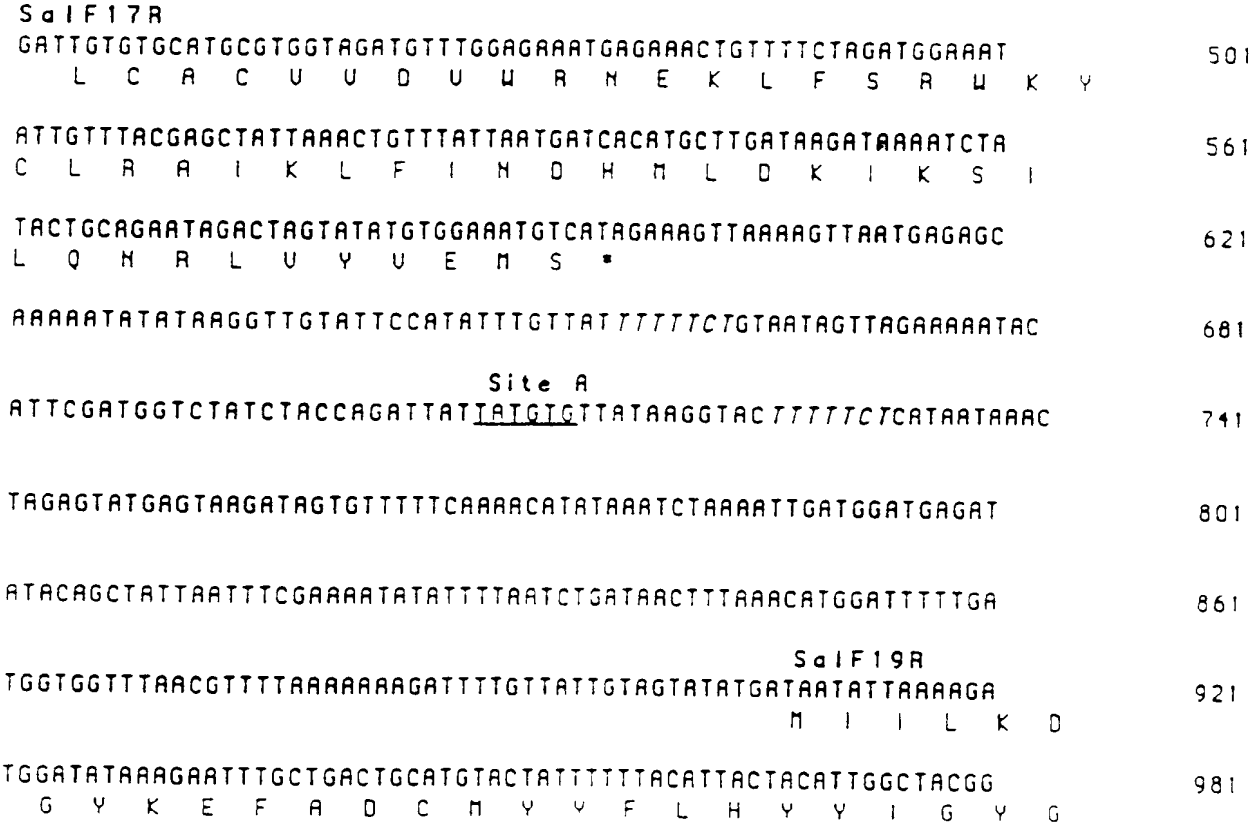

SalF17R
GATTGTGTGCATGCGTGGTAGATGTTTGGAGAAATGAGAAACTGTTTTCTAGATGGAAAT
L C A C U U D U W R N E K L F S R W K Y
501
ATTGTTTACGAGCTATTAAACTGTTTATTAATGATCACATGCTTGATAAGATAAAATCTA
C L R A I K L F I N D H M L D K I K S I
561
TACTGCAGAATAGACTAGTATATGTGGAAATGTCATAGAAAGTTAAAAGTTAATGAGAGC
L Q N R L U Y U E M S *
621
AAAAATATATAAGGTTGTATTCCATATTTGTTATTTTTTCTGTAATAGTTAGAAAAATAC
681
Site A
ATTCGATGGTCTATCTACCAGATTATTATGTGTTATAAGGTACTTTTTCTCATAATAAAC
741
TAGAGTATGAGTAAGATAGTGTTTTTCAAAACATATAAATCTAAAATTGATGGATGAGAT
801
ATACAGCTATTAATTTCGAAAATATATTTTAATCTGATAACTTTAAACATGGATTTTTGA
861
SalF19R
TGGTGGTTTAACGTTTTAAAAAAAGATTTTGTTATTGTAGTATATGATAATATTAAAAGA
M I I L K D
921
TGGATATAAAGAATTTGCTGACTGCATGTACTATTTTTTACATTACTACATTGGCTACGG
G Y K E F A D C M Y Y F L H Y Y I G Y G
981

Fig. 10.

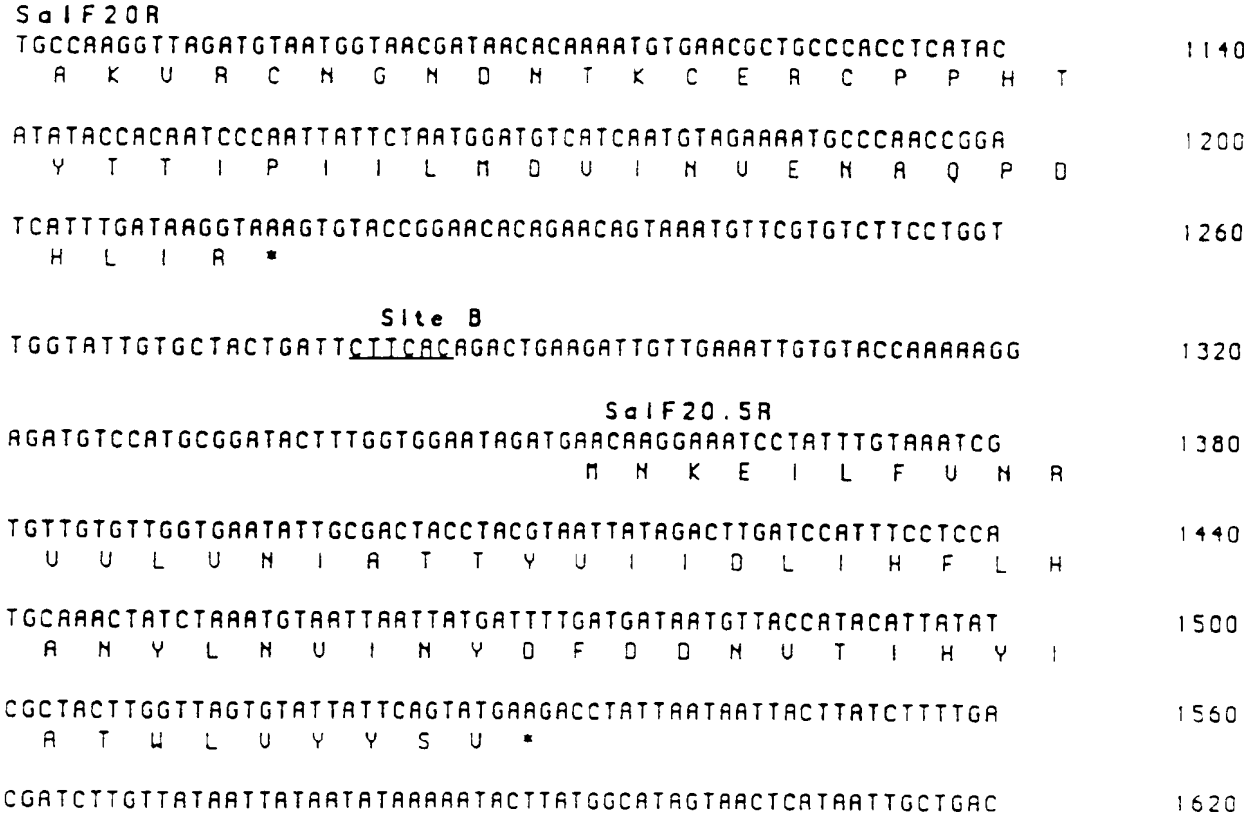

SalF20R
TGCCAAGGTTAGATGTAATGGTAACGATAACACAAAATGTGAACGCTGCCCACCTCATAC 1140
A K V R C N G N D N T K C E R C P P H T
ATATACCACAATCCCAATTATTCTAATGGATGTCATCAATGTAGAAAATGCCCAACCGGA 1200
Y T T I P I I L M D V I N V E N A Q P D
TCATTTGATAAGGTAAAGTGTACCGGAACACAGAACAGTAAATGTTCGTGTCTTCCTGGT 1260
H L I R *
Site B
TGGTATTGTGCTACTGATTCTTCACAGACTGAAGATTGTTGAAATTGTGTACCAAAAAGG 1320
SalF20.5R
AGATGTCCATGCGGATACTTTGGTGGAATAGATGAACAAGGAAATCCTATTTGTAAATCG 1380
M N K E I L F V N R
TGTTGTGTTGGTGAATATTGCGACTACCTACGTAATTATAGACTTGATCCATTTCCTCCA 1440
V V L V N I A T T Y V I I D L I H F L H
TGCAAACTATCTAAATGTAATTAATTATGATTTTGATGATAATGTTACCATACATTATAT 1500
A N Y L N V I N Y D F D D N V T I H Y I
CGCTACTTGGTTAGTGTATTATTCAGTATGAAGACCTATTAATAATTACTTATCTTTTGA 1560
A T W L V Y Y S V *
CGATCTTGTTATAATTATAATATAAAAATACTTATGGCATAGTAACTCATAATTGCTGAC 1620

Fig. 11.

Gap Weight.	3.000	Average Match.	0.540
Length Weight.	0.100	Average Mismatch.	-0.396
Quality.	113.0	Length.	199
Ratio.	0.608	Gaps.	4
Percent Similarity.	56.354	Percent Identity.	29.282

The top sequence is SalG2R and the bottom sequence is the yeast GK

Start of G2R

```
  1 MSGIVKSIILSGPSGLGKTAIAKRLMGIY...LDLWCPIPLDFLVLMERE  47
    ..:..::::: ::... :.:.. :   ..... .       : .
  1 ....SRPIVISGPSGTGKSTLLKKLFAEYPDSFGFSVSSTTRTPRAGEVN  46

 48 GVDYHYVNREAIWKGIAAGNFLEHTEFLGNIYGTSKTAVNTAAINNRICV  97
    : ::...:. ..... : ...:.: ..: ::.::.. ..:. .. .....:.
 47 GKDYNFVSVDEFKSMIKNNEFIEWAQFSGNYYGSTVASVKQVSKSGKTCI  96

 98 MDLNIDGVRSLKN.TYLMPYSVYIRPTSLKNVETKLRCRNTEANDEIHRR 146
    .:.....:.:.:. . : .  ..: :.:.. ....: .:.:.....:.:
 97 LDIDMQGVKSVKAIPELNARFLFIAPPSVEDLKKRLEGRGTETEESINKR 146

147 VILAKTDMDEANEAGLFDTIIIEDDVNLAYSKLIQ.ILQDRIRMYFNTN 194
    .  :.... : :.:  :.:.:.:.. :::.: . :....
147 LSARQAELAYA.ETGAHDKVIVNDDLDKAYKELKDFIFAEK..... 186
```

Fig.12.


HindB3R
CTAAGAACACGTATACGGCAGCAGCTTCCTTTATACTCTCATCTTTTACCAACACAAAGG 507
L R T R I R Q Q L P L Y S H L L P T Q R
GTGGATATTTGTTCATTGGAGTTGATAATAATACACACAAAGTAATTGGATTCACGGTGG 567
V D I C S L E L I I I H T K *
GTCATGACTACCTCAGACTGGTAGAGAATGATATAGAAAAGCATATCAAAAGACTTCGTG 627
TTGTGCATTTCTGTGAGAAGAAAGAGGACATCAAGTACACGTGTCGATTCATCAAGGTAT 687
Site D
ATAAACCTGGGGATGAGGCTACCTCGACATACGTGTGCGCTATCAAAGTGGAAAGATGCT 747
GTTGTGCTGTGTTTGCAGATTGGCCAGAATCATGGTATATGGATACTAATGGTATCAAGA 807
AGTATTCTCCAGATGAATGGGTGTCACATATAAAATTTTAATTAATGTAATAGAGAACAA 867
ATAATAAGGTTGTAATATCATATAGACAATAACTAACAATTAATTAGTAACTGTTATCTC 927
TTTTTTAACTAACCAACTAACTATATACCTATTAATACATCGTAATTATAGTTCTTAACA 987
TCTATTAATCATTAATTCGCTTCTTTAATTTTTTTATAAACTAACATTGTTAATTGAAAAG 1047
HindB4R
GGATAACATGTTACAGAATATAAATTATATATGGATTTTTTTAAAAAGGAAATACTTGAC 1107
M D F F K K E I L D
TGGAGTATATATTTATCTCTTCATTATATAGCACGCGTGTTTTCCAATTTTTCCACATCC 1167
W S I Y L S L H Y I A R V F S N F S T S

Fig.13.

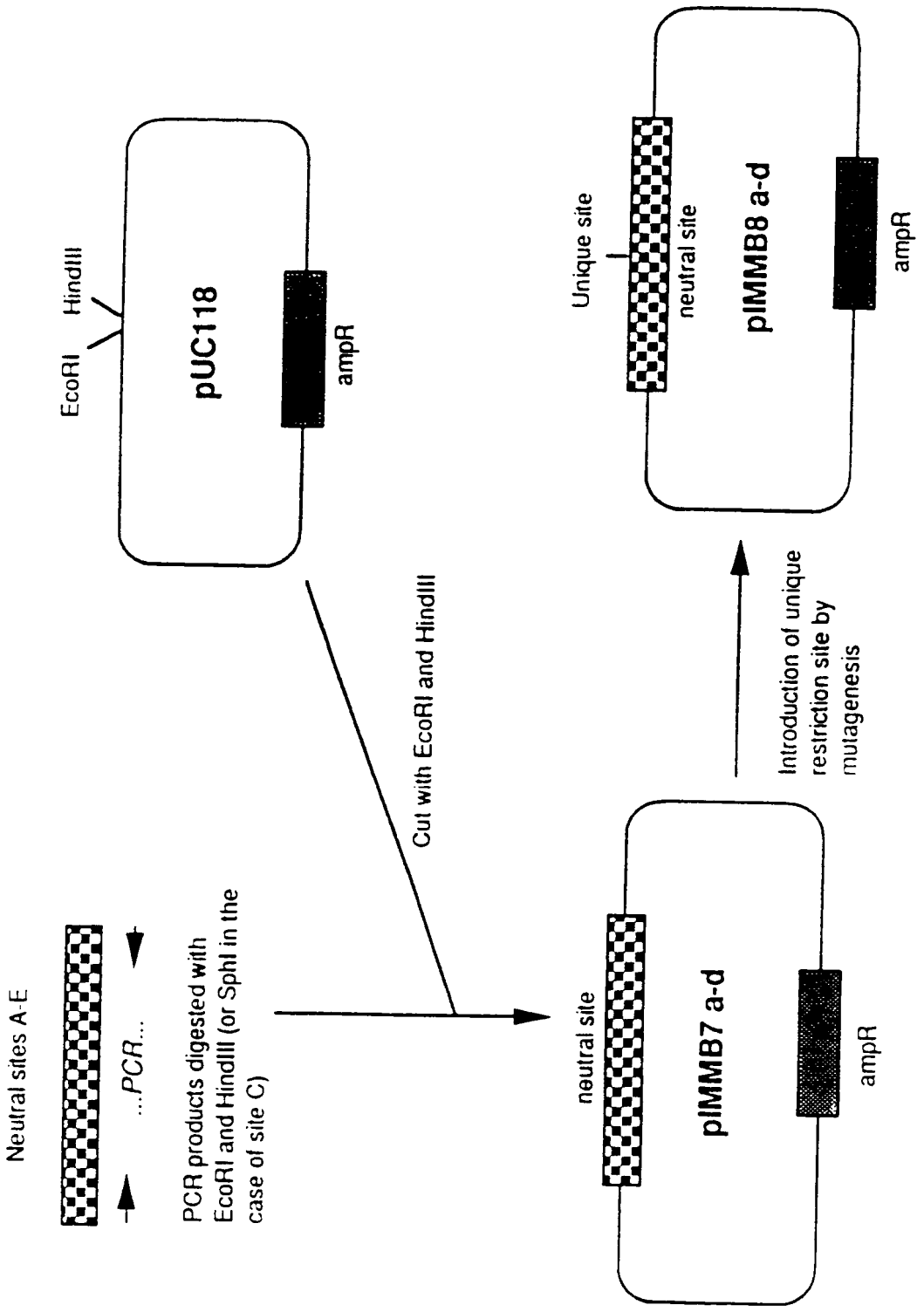
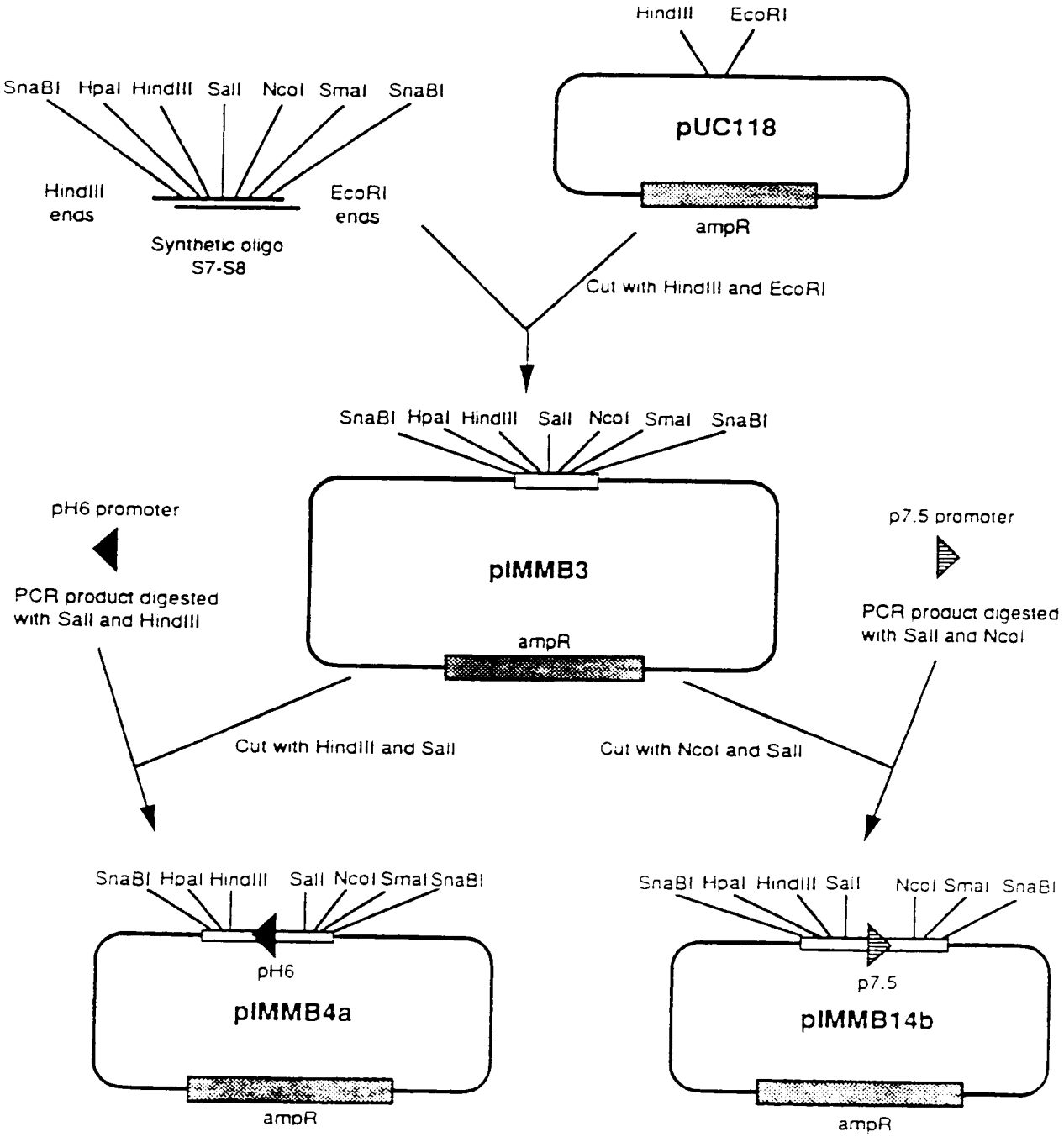

Neutral sites A-E
...PCR...
PCR products digested with EcoRI and HindIII (or SphI in the case of site C)
EcoRI
HindIII
pUC118
ampR
Cut with EcoRI and HindIII
neutral site
pIMMB7 a-d
ampR
Introduction of unique restriction site by mutagenesis
Unique site
neutral site
pIMMB8 a-d
ampR

Fig.14.


HindIII EcoRI
SnaBI HpaI HindIII SalI NcoI SmaI SnaBI
pUC118
HindIII ends
EcoRI ends
Synthetic oligo S7-S8
ampR
Cut with HindIII and EcoRI
SnaBI HpaI HindIII SalI NcoI SmaI SnaBI
pH6 promoter
pIMMB3
p7.5 promoter
PCR product digested with SalI and HindIII
ampR
PCR product digested with SalI and NcoI
Cut with HindIII and SalI
Cut with NcoI and SalI
SnaBI HpaI HindIII SalI NcoI SmaI SnaBI
pH6
pIMMB4a
ampR
SnaBI HpaI HindIII SalI NcoI SmaI SnaBI
p7.5
pIMMB14b
ampR

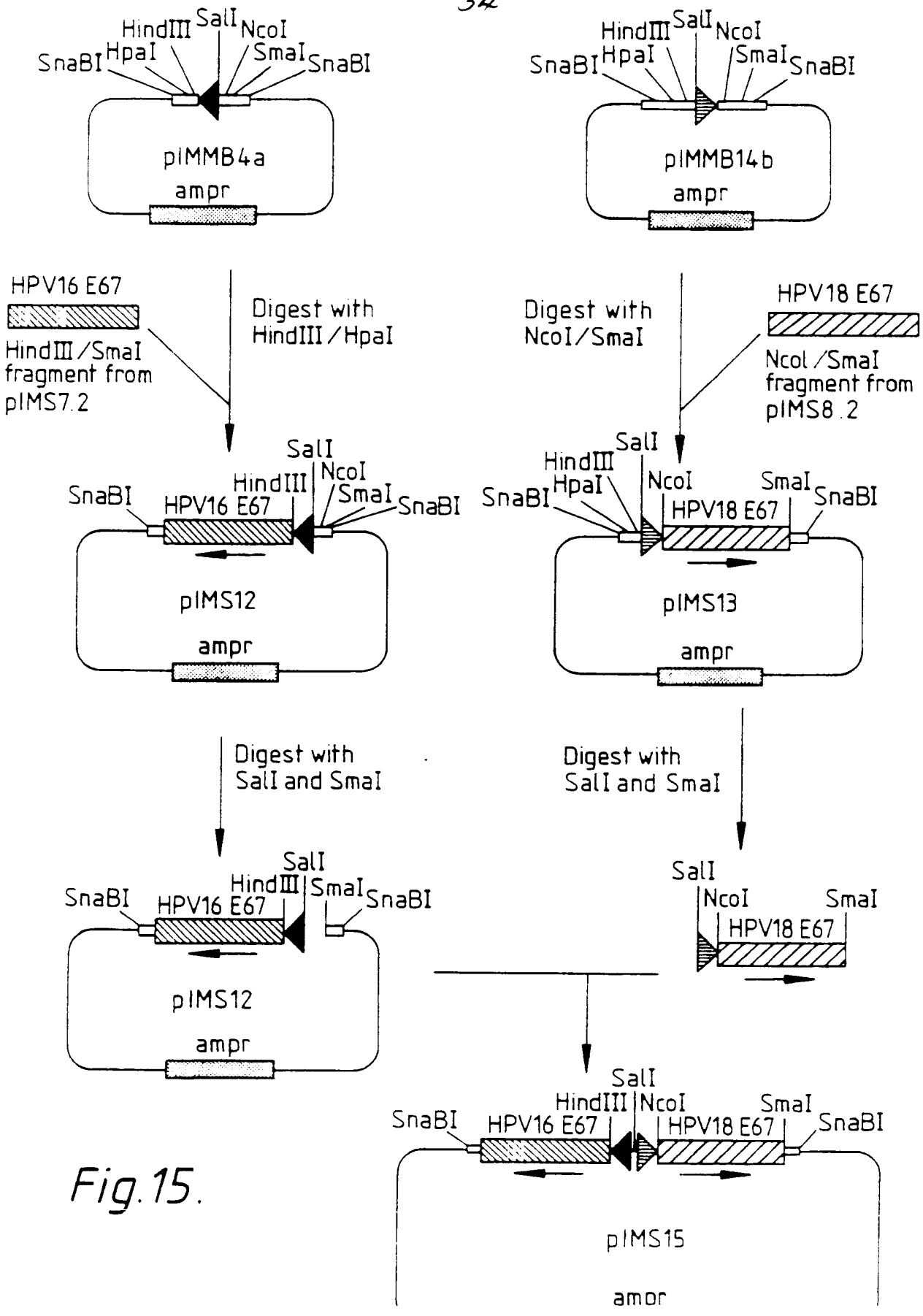

HindIII
SalI
NcoI
SnaBI
HpaI
SmaI
SnaBI
pIMMB4a
ampr
pIMMB14b
HPV16 E67
HindIII/SmaI fragment from pIMS7.2
Digest with HindIII/HpaI
HPV18 E67
Digest with NcoI/SmaI
NcoI/SmaI fragment from pIMS8.2
pIMS12
pIMS13
Digest with SalI and SmaI
pIMS15
amor

Fig. 15.

Fig. 16.

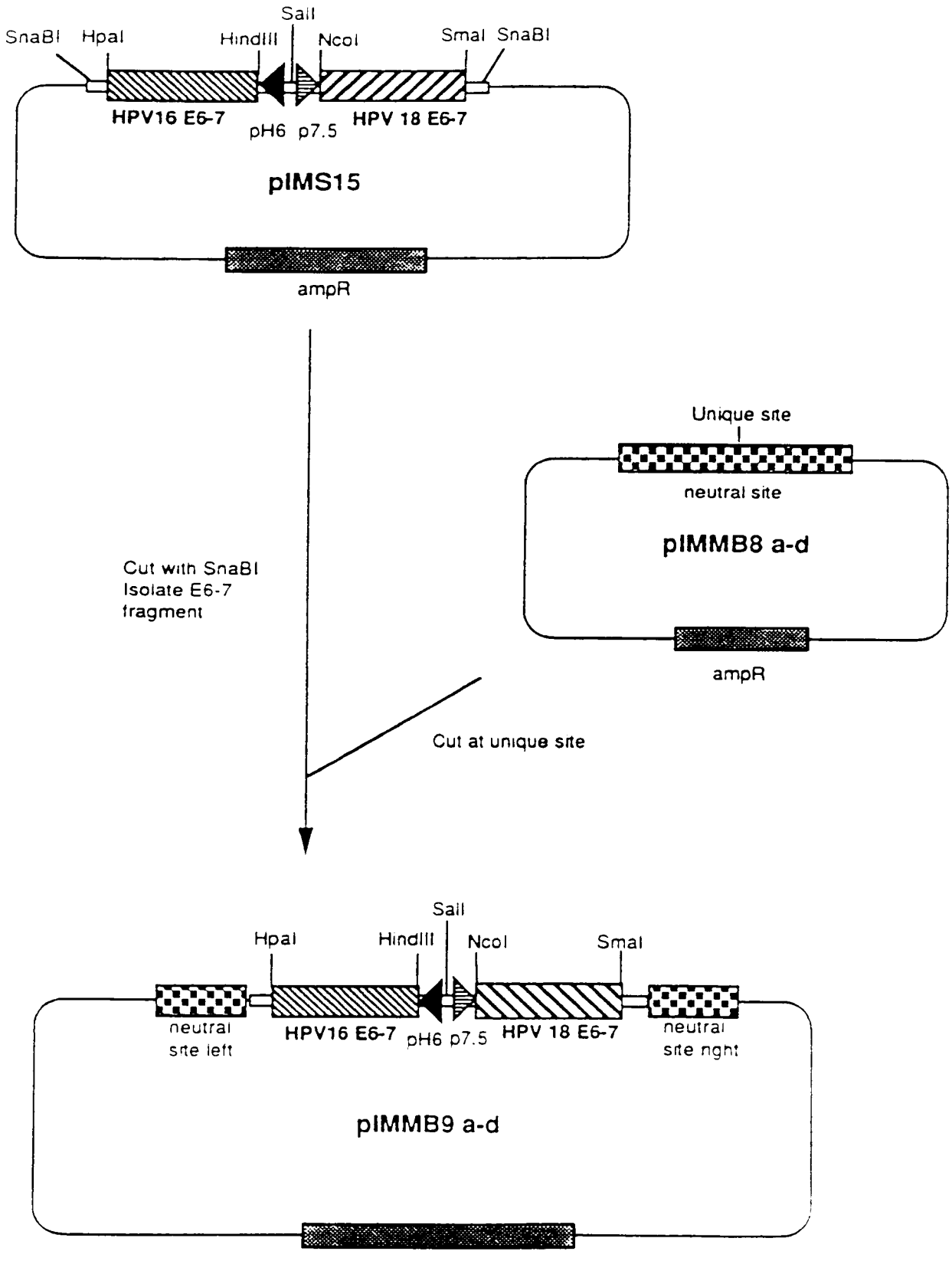

SnaBI
HpaI
HindIII
SalI
NcoI
SmaI
SnaBI
HPV16 E6-7
pH6 p7.5
HPV 18 E6-7
pIMS15
ampR
Cut with SnaBI
Isolate E6-7
fragment
Unique site
neutral site
pIMMB8 a-d
ampR
Cut at unique site
SalI
HpaI
HindIII
NcoI
SmaI
neutral
site left
HPV16 E6-7
pH6 p7.5
HPV 18 E6-7
neutral
site right
pIMMB9 a-d

Fig. 17.

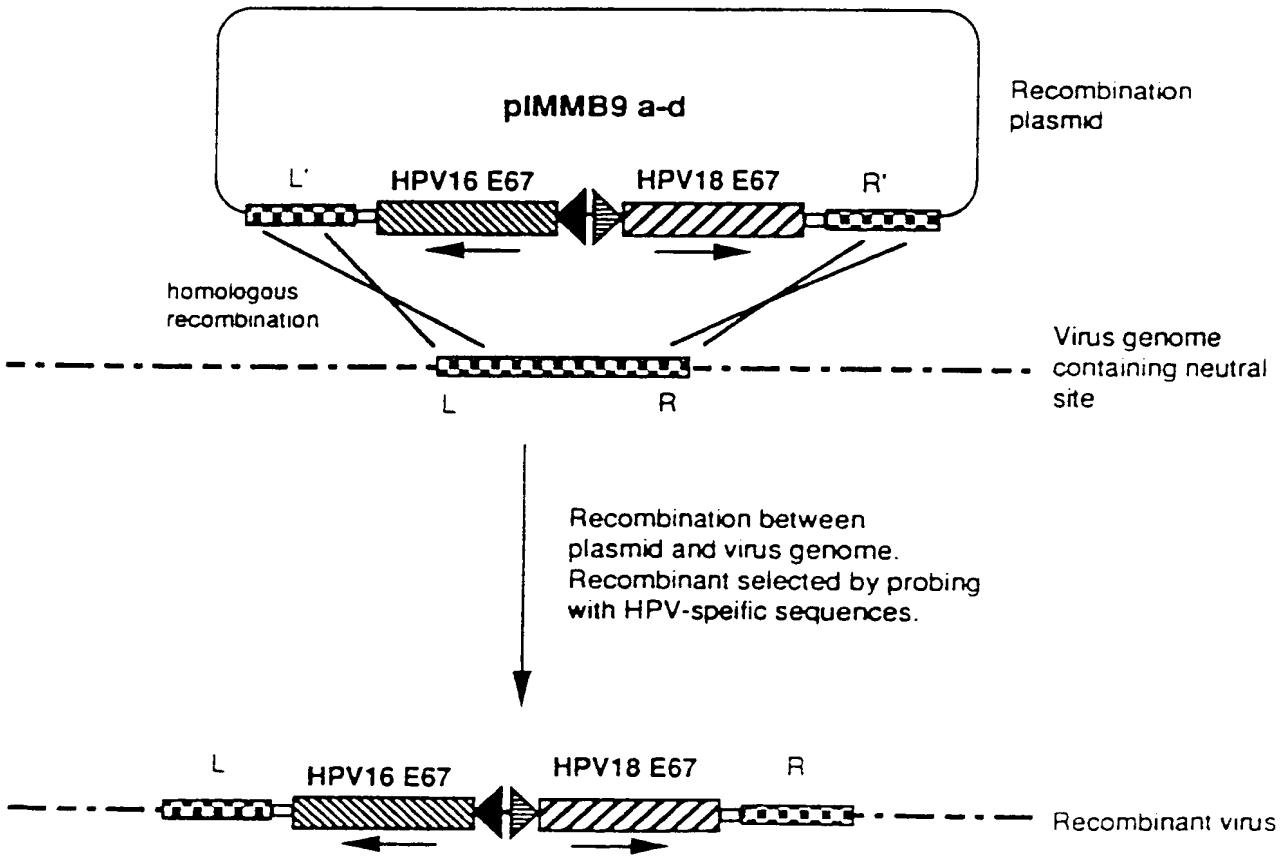

pIMMB9 a-d
Recombination plasmid
L'
HPV16 E67
HPV18 E67
R'
homologous recombination
Virus genome containing neutral site
L
R
Recombination between plasmid and virus genome. Recombinant selected by probing with HPV-speific sequences.
L
HPV16 E67
HPV18 E67
R
Recombinant virus

Fig. 18.

MB03	TCAGGATCCCACATGAGCGAAAAATACATCG
MB07	TCAAAGCTTATTACGATACAAACTTAACGGA
MB09	TCAGTTAACATAAAAAGAACAACGCCCGGCAG
MB14	TCAAGGGCCTCTATATAGTAATACCAATACTC
MB15	TCAGTCGACTTACAAACAACTAGGAAATTGG
MB16	TCAGAATTCTATGTACAGAGGTCTATTAGGC
MB17	TCAAAGCTTGTATGAGGTGGGCAGCGTTCAC
MB18	TCAGAATTCTTAATTATATTGTCGGCCGTGG
MB19	TCAGCATGCATGATCCGTTAGCTTTGGGCTC
MB22	TCAGAATTCGAAGCTCTAGAGTATCTTAGCG
MB23	TCAAAGCTTTCCTGTATTATATGGGATGTGG
MB24	TCAGAATTCATTGATGGATGAGATATACAGC
MB25	TCAAAGCTTTCACAAAATCG
MB26	TCAGAATTCCACGTATACGGCAGCAGCTTCC
MB27	TCAAAGCTTTGTTCTACGTCCATTTTCAAGC
MB32	TCAGTCGACATACCAATACTCAAGACTACGA
MB33	TCACCATGGATTGCTATTGATTGAGTACTGTTC
MB35	AGTACCTTATAATACGTAATAATCTGGTAG
MB36	AATCTTCAGTCTGTTAACAATCAGTAGCAC
MB37	TACAACGAAGCTAGGCCTCAACCATTTTAA
MB38	TTTGATAGCGCATACGTATGTCGAGGTAGC
MB39	AATAGGTATATAGTTAACTGGTTAGTTAAA
S01	ATCCCATGGCGCGCTTTGAGGATCCAAC
S02	TCACCCGGGTTACTGCTGGGATGCACACCAC
S05	ATCCCATGGACCAAAAGAGAACTGCAATGTTTC
S06	TCACCCGGGTTATGGTTTCTGAGAACAGATG
S07	AGCTCTACGTAGTTAACAAGCTTGTCGACCCATGGCCCGGGTACGTA
S08	AATTTACGTACCCGGGCCATGGGTCGACAAGCTTGTTAACTACGTAG
S20	GAAACCCAGCTGGGAATCATGCATGG
S21	GAAACACAAGTAGGAATATTAAGTATG
S22	GATCTCTACGGTTATGGGCAATTAAATGAC
S23	GACCTTCTAGGTCACGGGCAATTAAGCGAC
S24	ATGTATAGATTTCTACAGTAGAATCAGAGAATTAAG

Fig.19.

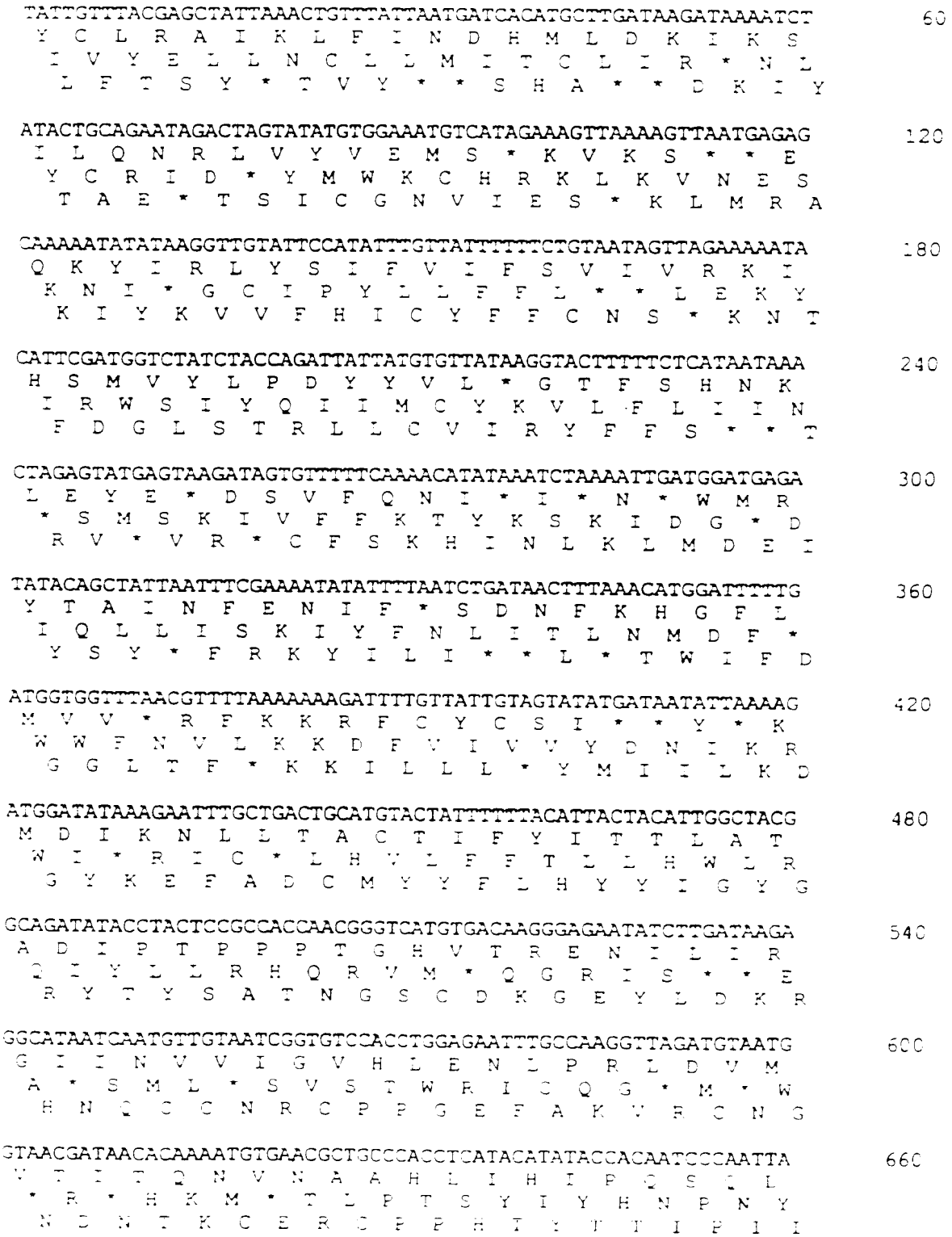

TATTGTTTACGAGCTATTAAACTGTTTATTAATGATCACATGCTTGATAAGATAAAATCT 60
Y C L R A I K L F I N D H M L D K I K S
I V Y E L L N C L L M I T C L I R * N L
L F T S Y * T V Y * * S H A * * D K I Y
ATACTGCAGAATAGACTAGTATATGTGGAAATGTCATAGAAAGTTAAAAGTTAATGAGAG 120
I L Q N R L V Y V E M S * K V K S * * E
Y C R I D * Y M W K C H R K L K V N E S
T A E * T S I C G N V I E S * K L M R A
CAAAAATATATAAGGTTGTATTCCATATTTGTTATTTTTTCTGTAATAGTTAGAAAAATA 180
Q K Y I R L Y S I F V I F S V I V R K I
K N I * G C I P Y L L F F L * * L E K Y
K I Y K V V F H I C Y F F C N S * K N T
CATTCGATGGTCTATCTACCAGATTATTATGTGTTATAAGGTACTTTTTCTCATAATAAA 240
H S M V Y L P D Y Y V L * G T F S H N K
I R W S I Y Q I I M C Y K V L F L I I N
F D G L S T R L L C V I R Y F F S * * T
CTAGAGTATGAGTAAGATAGTGTTTTTCAAAACATATAAATCTAAAATTGATGGATGAGA 300
L E Y E * D S V F Q N I * I * N * W M R
* S M S K I V F F K T Y K S K I D G * D
R V * V R * C F S K H I N L K L M D E I
TATACAGCTATTAATTTCGAAAATATATTTTAATCTGATAACTTTAAACATGGATTTTTG 360
Y T A I N F E N I F * S D N F K H G F L
I Q L L I S K I Y F N L I T L N M D F *
Y S Y * F R K Y I L I * * L * T W I F D
ATGGTGGTTTAACGTTTTAAAAAAAGATTTTGTTATTGTAGTATATGATAATATTAAAAG 420
M V V * R F K K R F C Y C S I * * Y * K
W W F N V L K K D F V I V V Y D N I K R
G G L T F * K K I L L L * Y M I I L K D
ATGGATATAAAGAATTTGCTGACTGCATGTACTATTTTTTACATTACTACATTGGCTACG 480
M D I K N L L T A C T I F Y I T T L A T
W I * R I C * L H V L F F T L L H W L R
G Y K E F A D C M Y Y F L H Y Y I G Y G
GCAGATATACCTACTCCGCCACCAACGGGTCATGTGACAAGGGAGAATATCTTGATAAGA 540
A D I P T P P P T G H V T R E N I L I R
Q I Y L L R H Q R V M * Q G R I S * * E
R Y T Y S A T N G S C D K G E Y L D K R
GGCATAATCAATGTTGTAATCGGTGTCCACCTGGAGAATTTGCCAAGGTTAGATGTAATG 600
G I I N V V I G V H L E N L P R L D V M
A * S M L * S V S T W R I C Q G * M * W
H N Q C C N R C P P G E F A K V R C N G
GTAACGATAACACAAAATGTGAACGCTGCCCACCTCATACATATACCACAATCCCAATTA 660
V T I T Q N V N A A H L I H I P Q S Q L
* R * H K M * T L P T S Y I Y H N P N Y
N D N T K C E R C P P H T Y T T I P I I

Fig.19(Cont).

```
TTCTAATGGATGTCATCAATGTAGAAAATGCCCAACCGGATCATTTGATAAGGTAAAGTG        720
 F  *  W  M  S  S  M  *  K  M  P  N  R  I  I  *  *  G  K  V
  S  N  G  C  H  Q  C  R  K  C  P  T  G  S  F  D  K  V  K  C
   L  M  D  V  I  N  V  E  N  A  Q  P  D  H  L  I  R  *  S  V

TACCGGAACACAGAACAGTAAATGTTCGTGTCTTCCTGGTTGGTATTGTGCTACTGATTC        780
 Y  R  N  T  E  Q  *  M  F  V  S  S  W  L  V  L  C  Y  *  F
  T  G  T  Q  N  S  K  C  S  C  L  P  G  W  Y  C  A  T  D  S
   P  E  H  R  T  V  N  V  R  V  F  L  V  G  I  V  L  L  I  L

TTCACAGACTGAAGATTGTTGAAATTGTGTACCAAAAAGGAGATGTCCATGCGGATACTT        840
 F  T  D  *  R  L  L  K  L  C  T  K  K  E  M  S  M  R  I  L
  S  Q  T  E  D  C  *  N  C  V  P  K  R  R  C  P  C  G  Y  F
   H  R  L  K  I  V  E  I  V  Y  Q  K  G  D  V  H  A  D  T  L

TGGTGGAATAGATGAACAAGGAAATCCTATTTGTAAATCGTGTTGTGTTGGTGAATATTG        900
 W  W  N  R  *  T  R  K  S  Y  L  *  I  V  L  C  W  *  I  L
  G  G  I  D  E  Q  G  N  P  I  C  K  S  C  C  V  G  E  Y  C
   V  E  *  M  N  K  E  I  L  F  V  N  R  V  V  L  V  N  I  A

CGACTACCTACGTAATTATAGACTTGATCCATTTCCTCCATGCAAACTATCTAAATGTAA        960
 R  L  P  T  *  L  *  T  *  S  I  S  S  M  Q  T  I  *  M  *
  D  Y  L  R  N  Y  R  L  D  P  F  P  P  C  K  L  S  K  C  N
   T  T  Y  V  I  I  D  L  I  H  F  L  H  A  N  Y  L  N  V  I

TTAATTATGATTTTGATGATAATGTTACCATACATTATATCGCTACTTGGTTAGTGTATT        1020
 L  I  M  I  L  M  I  M  L  P  Y  I  I  S  L  L  G  *  C  I
  *  L  *  F  *  *  *  C  Y  H  T  L  Y  R  Y  L  V  S  V  L
   N  Y  D  F  D  D  N  V  T  I  H  Y  I  A  T  W  L  V  Y  Y

ATTCAGTATGAAGACCTATTAATAATTACTTATCTTTTGACGATCTTGTTATAATTATAA        1080
 I  Q  Y  E  D  L  L  I  I  T  Y  L  L  T  I  L  L  *  L  *
  F  S  M  K  T  Y  *  *  L  L  I  F  *  R  S  C  Y  N  Y  N
   S  V  *  R  P  I  N  N  Y  L  S  F  D  D  L  V  I  I  I  I

TATAAAAATACTTATGGCATAGTAACTCATAATTGCTGACGCGATAAATTCGTAATAATC        1140
 Y  K  N  T  Y  G  I  V  T  H  N  C  *  R  D  K  F  V  I  I
  I  K  I  L  M  A  *  *  L  I  I  A  D  A  I  N  S  *  *  S
   *  K  Y  L  W  H  S  N  S  *  L  L  T  R  *  I  R  N  N  L

TGTTTTGTTCAAATTTTTATAAGGAATCTACAGGCATAAAAAATAAAAAATATAATTTATAA        1200
 C  F  V  Q  I  F  I  R  N  L  Q  A  *  K  *  K  Y  N  L  *
  V  L  F  K  F  L  *  G  I  Y  R  H  K  N  K  N  I  I  Y  N
   F  C  S  N  F  Y  K  E  S  T  G  I  K  I  K  I  *  F  I  I

TATACTCTTACAGCGCGCCATCATGAATAACAGCAGTGAATTGATTGCTG                  1260
 Y  T  L  T  A  R  H  H  E  *  Q  Q  *  I  D  C
  I  L  L  Q  R  A  I  M  N  N  S  S  E  L  I  A
   Y  S  Y  S  A  P  S  *  I  T  A  V  N  *  L  L
```

Fig. 20.

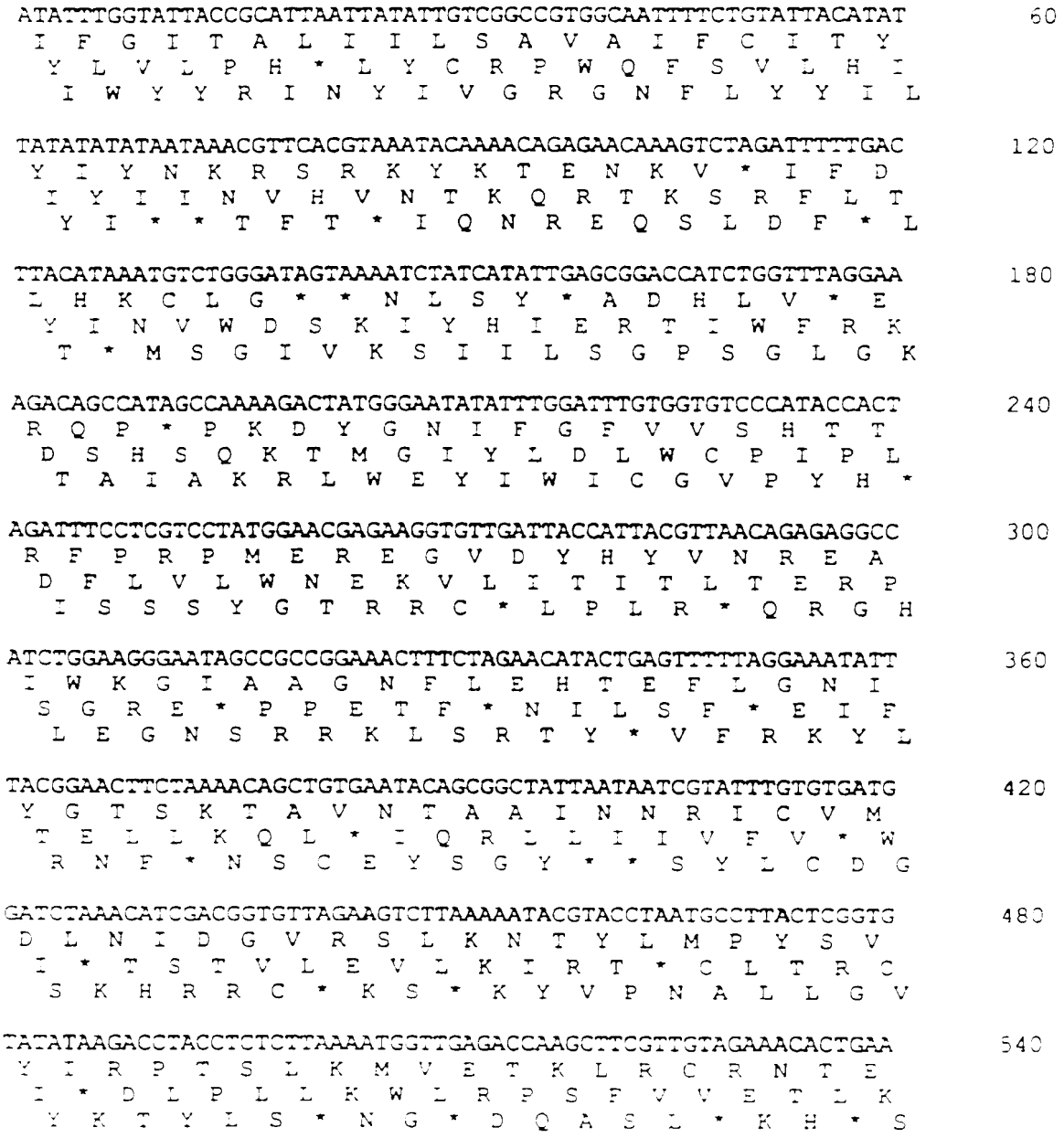

ATATTTGGTATTACCGCATTAATTATATTGTCGGCCGTGGCAATTTTCTGTATTACATAT 60
I F G I T A L I I L S A V A I F C I T Y
Y L V L P H * L Y C R P W Q F S V L H I
I W Y Y R I N Y I V G R G N F L Y Y I L
TATATATATAATAAACGTTCACGTAAATACAAAACAGAGAACAAAGTCTAGATTTTTGAC 120
Y I Y N K R S R K Y K T E N K V * I F D
I Y I I N V H V N T K Q R T K S R F L T
Y I * * T F T * I Q N R E Q S L D F * L
TTACATAAATGTCTGGGATAGTAAAATCTATCATATTGAGCGGACCATCTGGTTTAGGAA 180
L H K C L G * * N L S Y * A D H L V * E
Y I N V W D S K I Y H I E R T I W F R K
T * M S G I V K S I I L S G P S G L G K
AGACAGCCATAGCCAAAAGACTATGGGAATATATTTGGATTTGTGGTGTCCCATACCACT 240
R Q P * P K D Y G N I F G F V V S H T T
D S H S Q K T M G I Y L D L W C P I P L
T A I A K R L W E Y I W I C G V P Y H *
AGATTTCCTCGTCCTATGGAACGAGAAGGTGTTGATTACCATTACGTTAACAGAGAGGCC 300
R F P R P M E R E G V D Y H Y V N R E A
D F L V L W N E K V L I T I T L T E R P
I S S S Y G T R R C * L P L R * Q R G H
ATCTGGAAGGGAATAGCCGCCGGAAACTTTCTAGAACATACTGAGTTTTTAGGAAATATT 360
I W K G I A A G N F L E H T E F L G N I
S G R E * P P E T F * N I L S F * E I F
L E G N S R R K L S R T Y * V F R K Y L
TACGGAACTTCTAAAACAGCTGTGAATACAGCGGCTATTAATAATCGTATTTGTGTGATG 420
Y G T S K T A V N T A A I N N R I C V M
T E L L K Q L * I Q R L L I I V F V * W
R N F * N S C E Y S G Y * * S Y L C D G
GATCTAAACATCGACGGTGTTAGAAGTCTTAAAAATACGTACCTAATGCCTTACTCGGTG 480
D L N I D G V R S L K N T Y L M P Y S V
I * T S T V L E V L K I R T * C L T R C
S K H R R C * K S * K Y V P N A L L G V
TATATAAGACCTACCTCTCTTAAAATGGTTGAGACCAAGCTTCGTTGTAGAAACACTGAA 540
Y I R P T S L K M V E T K L R C R N T E
I * D L P L L K W L R P S F V V E T L K
Y K T Y L S * N G * D Q A S L * K H * S

Fig. 20 (Cont).

```
GCTAACGATGAGATTCATCGTCGCGTGATATTGGCAAAAACGGATATGGATGAGGCCAAC      600
 A  N  D  E  I  H  R  R  V  I  L  A  K  T  D  M  D  E  A  N
  L  T  M  R  F  I  V  A  *  Y  W  Q  K  R  I  W  M  R  P  T
   *  R  *  D  S  S  S  R  D  I  G  K  N  G  Y  G  *  G  Q  R

GAAGCAGGTCTATTCGACACTATTATCATTGAAGATGATGTGAATTTAGCATATAGTAAG      660
 E  A  G  L  F  D  T  I  I  I  E  D  D  V  N  L  A  Y  S  K
  K  Q  V  Y  S  T  L  L  S  L  K  M  M  *  I  *  H  I  V  S
   S  R  S  I  R  H  Y  Y  H  *  P  *  C  E  F  S  I  *  *  V

TTAATTCAGATACTACAGGACCGTATTAGAATGTATTTTAACACTAATTAGAGACTTAAG      720
 L  I  Q  I  L  Q  D  R  I  R  M  Y  F  N  T  N  *  R  L  K
  *  F  R  Y  Y  R  T  V  L  E  C  I  L  T  L  I  R  D  L  R
   N  S  D  T  T  G  P  Y  *  N  V  F  *  H  *  L  E  T  *  D

ACTTAAAACTTGATAATTAATAATATAACTCGTTTTTATATGTGGCTATTTCAACGTCTA      780
 T  *  N  L  I  I  N  N  I  T  R  F  Y  M  W  L  F  Q  R  L
  L  K  T  *  *  L  I  I  *  L  V  F  I  C  G  Y  F  N  V  *
   L  K  L  D  N  *  *  Y  N  S  F  L  Y  V  A  I  S  T  S  N

ATGTATTAGTTAAATATTAAAACTTACCACGTAAAACTTAAAATTTAAAATGATATTTCA      840
 M  Y  *  L  N  I  K  T  Y  H  V  K  L  K  I  *  N  D  I  S
  C  I  S  *  I  L  K  L  T  T  *  N  L  K  F  K  M  I  F  H
   V  L  V  K  Y  *  N  L  P  R  K  T  *  N  L  K  *  Y  F  I

TTGACAGATAGATCACACATTATGAACTTTCAAGGACTTGTGTTAACTGACAATTGCAAA      900
 L  T  D  R  S  H  I  M  N  F  Q  G  L  V  L  T  D  N  C  K
  *  Q  I  D  H  T  L  *  T  F  K  D  L  C  *  L  T  I  A  K
   D  R  *  I  T  H  Y  E  L  S  R  T  C  V  N  *  Q  L  Q  K

AATCAATGGGTCGTTGGACCATTAATAGGAAAAGGTGGATTCGGTAGTATTTATACTACT      960
 N  Q  W  V  V  G  P  L  I  G  K  G  G  F  G  S  I  Y  T  T
  I  N  G  S  L  D  H  *  *  E  K  V  D  S  V  V  F  I  L  L
   S  M  G  R  W  T  I  N  R  K  R  W  I  R  *  Y  L  Y  Y  *

AATGACAATAATTATGTAGTAAAAATAGAGCCCAAAGCTA                         1020
 N  D  N  N  Y  V  V  K  I  E  P  K  A
  M  T  I  I  M  *  *  K  *  S  P  K  L
   *  Q  *  L  C  S  K  N  R  A  Q  S  S
```

Fig. 21.

```
ACCATCGAGGTAACCACCTCTCTGGAAGACAGCGTGAATAATGTACTCATGAAACGTTTG          60
 T  I  E  V  T  T  S  L  E  D  S  V  N  N  V  L  M  K  R  L
  P  S  R  *  P  P  L  W  K  T  A  *  I  M  Y  S  *  N  V  W
   H  R  G  N  H  L  S  G  R  Q  R  E  *  C  T  H  E  T  F  G

GAAACTATACGCCATATGTGGTCTGTCGTATATGATCATTTTGATATTGTGAATGGTAAA         120
 E  T  I  R  H  M  W  S  V  V  Y  D  H  F  D  I  V  N  G  K
  K  L  Y  A  I  C  G  L  S  Y  M  I  I  L  I  L  *  M  V  K
   N  Y  T  P  Y  V  V  C  R  I  *  S  F  *  Y  C  E  W  *  R

GAATGCTGTTATGTGCATACGCATTTGTCTAATCAAAATCTTATACCGAGTACTGTAAAA         180
 E  C  C  Y  V  H  T  H  L  S  N  Q  N  L  I  P  S  T  V  K
  N  A  V  M  C  I  R  I  C  L  I  K  I  L  Y  R  V  L  *  K
   M  L  L  C  A  Y  A  F  V  *  S  K  S  Y  T  E  Y  C  K  N

ACAAATTTGTACATGAAGACTATGGGATCATGCATTCAAATGGATTCCATGGAAGCTCTA         240
 T  N  L  Y  M  K  T  M  G  S  C  I  Q  M  D  S  M  E  A  L
  Q  I  C  T  *  R  L  W  D  H  A  F  K  W  I  P  W  K  L  *
   K  F  V  H  E  D  Y  G  I  M  H  S  N  G  F  H  G  S  S  R

GAGTATCTTAGCGAACTGAAGGAATCAGGTGGATGGAGTCCCAGACCAGAAATGCAGGAA         300
 E  Y  L  S  E  L  K  E  S  G  G  W  S  P  R  P  E  M  Q  E
  S  I  L  A  N  *  R  N  Q  V  D  G  V  P  D  Q  K  C  R  N
   V  S  *  R  T  E  G  I  R  W  M  E  S  Q  T  R  N  A  G  I

TTTGAATATCCAGATGGAGTGGAAGACACTGAATCAATTGAGAGATTGGTAGAGGAGTTC         360
 F  E  Y  P  D  G  V  E  D  T  E  S  I  E  R  L  V  E  E  F
  L  N  I  Q  M  E  W  K  T  L  N  Q  L  R  D  W  *  R  S  S
   *  I  S  R  W  S  G  R  H  *  I  N  *  E  I  G  R  G  V  L

TTCAATAGATCAGAACTTCAGGCTGGTGAATCAGTCAAATTTGGTAATTCTATTAATGTT         420
 F  N  R  S  E  L  Q  A  G  E  S  V  K  F  G  N  S  I  N  V
  S  I  D  Q  N  F  R  L  V  N  Q  S  N  L  V  I  L  L  M  L
   Q  *  I  R  T  S  G  W  *  I  S  Q  I  W  *  F  Y  *  C  *

AAACATACATCTGTTTCAGCTAAGCAACTAAGAACACGTATACGGCAGCAGCTTCCTTTA         480
 K  H  T  S  V  S  A  K  Q  L  R  T  R  I  R  Q  Q  L  P  L
  N  I  H  L  F  Q  L  S  N  *  E  H  V  Y  G  S  S  F  L  Y
   T  Y  I  C  F  S  *  A  T  K  N  T  Y  T  A  A  A  S  F  I

TACTCTCATCTTTTACCAACACAAAGGGTGGATATTTGTTCATTGGAGTTGATAATAATA         540
 Y  S  H  L  L  P  T  Q  R  V  D  I  C  S  L  E  L  I  I  I
  T  L  I  F  Y  Q  H  K  G  W  I  F  V  H  W  S  *  *  *  Y
   L  S  S  F  T  N  T  K  G  G  Y  L  F  I  G  V  D  N  N  T

CACACAAAGTAATTGGATTCACGGTGGGTCATGACTACCTCAGACTGGTAGAGAATGATA         600
 H  T  K  *  L  D  S  R  W  V  M  T  T  S  D  W  *  R  M  I
  T  Q  S  N  W  I  H  G  G  S  *  L  P  Q  T  G  R  E  *  Y
   H  K  V  I  G  F  T  V  G  H  D  Y  L  R  L  V  E  N  D  I

TAGAAAAGCATATCAAAAGACTTCGTGTTGTGCATTTCTGTGAGAAGAAAGAGGACATCA         660
 *  K  S  I  S  K  D  F  V  L  C  I  S  V  R  R  K  R  T  S
  R  K  A  Y  Q  K  T  S  C  C  A  F  L  *  E  E  R  G  H  Q
   E  K  H  I  K  R  L  R  V  V  H  F  C  E  K  K  E  D  I  K
```

Fig. 21 (Cont 1).

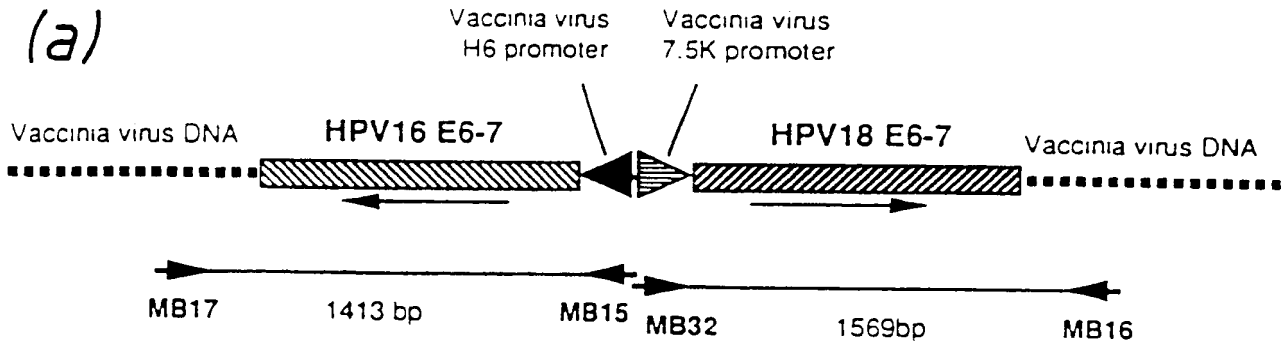
```
AGTACACGTGTCGATTCATCAAGGTATATAAACCTGGGGATGAGGCTACCTCGACATACG        720
 S  T  R  V  D  S  S  R  Y  I  N  L  G  M  R  L  P  R  H  T
  V  H  V  S  I  H  Q  G  I  *  T  W  G  *  G  Y  L  D  I  R
   Y  T  C  R  F  I  K  V  Y  K  P  G  D  E  A  T  S  T  Y  V

TGTGCGCTATCAAAGTGGAAAGATGCTGTTGTGCTGTGTTTGCAGATTGGCCAGAATCAT        780
 C  A  L  S  K  W  K  D  A  V  V  L  C  L  Q  I  G  Q  N  H
  V  R  Y  Q  S  G  K  M  L  L  C  C  V  C  R  L  A  R  I  M
   C  A  I  K  V  E  R  C  C  C  A  V  F  A  D  W  P  E  S  W

GGTATATGGATACTAATGGTATCAAGAAGTATTCTCCAGATGAATGGGTGTCACATATAA        840
 G  I  W  I  L  M  V  S  R  S  I  L  Q  M  N  G  C  H  I  *
  V  Y  G  Y  *  W  Y  Q  E  V  F  S  R  *  M  G  V  T  Y  K
   Y  M  D  T  N  G  I  K  K  Y  S  P  D  E  W  V  S  H  I  K

AATTTTAATTAATGTAATAGAGAACAAATAATAAGGTTGTAATATCATATAGACAATAAC        900
 N  F  N  *  C  N  R  E  Q  I  I  R  L  *  Y  H  I  D  N  N
  I  L  I  N  V  I  E  N  K  *  *  G  C  N  I  I  *  T  I  T
   F  *  L  M  *  *  R  T  N  N  K  V  V  I  S  Y  R  Q  *  L

TAACAATTAATTAGTAACTGTTATCTCTTTTTTAACTAACCAACTAACTATATACCTATT        960
 *  Q  L  I  S  N  C  Y  L  F  F  N  *  P  T  N  Y  I  P  I
  N  N  *  L  V  T  V  I  S  F  L  T  N  Q  L  T  I  Y  L  L
   T  I  N  *  *  L  L  S  L  F  *  L  T  N  *  L  Y  T  Y  *

AATACATCGTAATTATAGTTCTTAACATCTATTAATCATTAATTCGCTTCTTTAATTTTT       1020
 N  T  S  *  L  *  F  L  T  S  I  N  H  *  F  A  S  L  I  F
  I  H  R  N  Y  S  S  *  H  L  L  I  I  N  S  L  L  *  F  F
   Y  I  V  I  I  V  L  N  I  Y  *  S  L  I  R  F  F  N  F  L

TATAAACTAACATTGTTAATTGAAAAGGGATAACATGTTACAGAATATAAATTATATATG       1080
 Y  K  L  T  L  L  I  E  K  G  *  H  V  T  E  Y  K  L  Y  M
  I  N  *  H  C  *  L  K  R  D  N  M  L  Q  N  I  N  Y  I  W
   *  T  N  I  V  N  *  K  G  I  T  C  Y  R  I  *  I  I  Y  G

GATTTTTTTAAAAAGGAAATACTTGACTGGAGTATATATTTATCTCTTCATTATATAGCA       1140
 D  F  F  K  K  E  I  L  D  W  S  I  Y  L  S  L  H  Y  I  A
  I  F  L  K  R  K  Y  L  T  G  V  Y  I  Y  L  F  I  I  *  H
   F  F  *  K  G  N  T  *  L  E  Y  I  F  I  S  S  L  Y  S  T
```

Fig. 21(Cont 2).

```
CGCGTGTTTTCCAATTTTTCCACATCCCATATAATACAGGATTATAATCTCGTTCGAACA      1200
 R  V  F  S  N  F  S  T  S  H  I  I  Q  D  Y  N  L  V  R  T
  A  C  F  P  I  F  P  H  P  I  *  Y  R  I  I  I  S  F  E  H
   R  V  F  Q  F  F  H  I  P  Y  N  T  G  L  *  S  R  S  N  I

TACGAGAAAGTGGATAAAACAATAGTTGATTTTTTATCTAGGTTGCCAAATTTATTCCAT      1260
 Y  E  K  V  D  K  T  I  V  D  F  L  S  R  L  P  N  L  F  H
  T  R  K  W  I  K  Q  *  L  I  F  Y  L  G  C  Q  I  Y  S  I
   R  E  S  G  *  N  N  S  *  F  F  I  *  V  A  K  F  I  P  Y

ATTTTAGAATATGGGGAAAATATTCTACATATTTATTCTATGGATGATGCTAATACGAAT      1320
 I  L  E  Y  G  E  N  I  L  H  I  Y  S  M  D  D  A  N  T  N
  F  *  N  M  G  K  I  F  Y  I  F  I  L  W  M  M  L  I  R  I
   F  R  I  W  G  K  Y  S  T  Y  L  F  Y  G  *  C  *  Y  E  Y

ATTATAATTTTTTTTCTAGATAGAGTATTAAATATTAATAAGAACGGGTCATTTATACAC      1380
 I  I  I  F  F  L  D  R  V  L  N  I  N  K  N  G  S  F  I  H
  L  *  F  F  F  *  I  E  Y  *  I  L  I  R  T  G  H  L  Y  T
   Y  N  F  F  S  R  *  S  I  K  Y  *  *  E  R  V  I  Y  T  Q

AATCTCAGGTTATCATCATCCATTAATATAAAAGAATATGTATATCAATTAGTTAATAAT      1440
 N  L  R  L  S  S  S  I  N  I  K  E  Y  V  Y  Q  L  V  N  N
  I  S  G  Y  H  H  P  L  I  *  K  N  M  Y  I  N  *  L  I  M
   S  Q  V  I  I  I  H  *  Y  K  R  I  C  I  S  I  S  *  *  *

GATCATCCAGATAATAGGATAAGACTAATGCTTGAAAATGGACGTAGAACAAGACATTTT      1500
 D  H  P  D  N  R  I  R  L  M  L  E  N  G  R  R  T  R  H  F
  I  I  Q  I  I  G  *  D  *  C  L  K  M  D  V  E  Q  D  I  I
   S  S  R  *  *  D  K  T  N  A  *  K  W  T  *  N  K  T  F  F
```

Fig. 22.

(a)



The sequences present at the 5' end of PCR primers M15 and MB32 overlap by 6 nucleotides

(b)

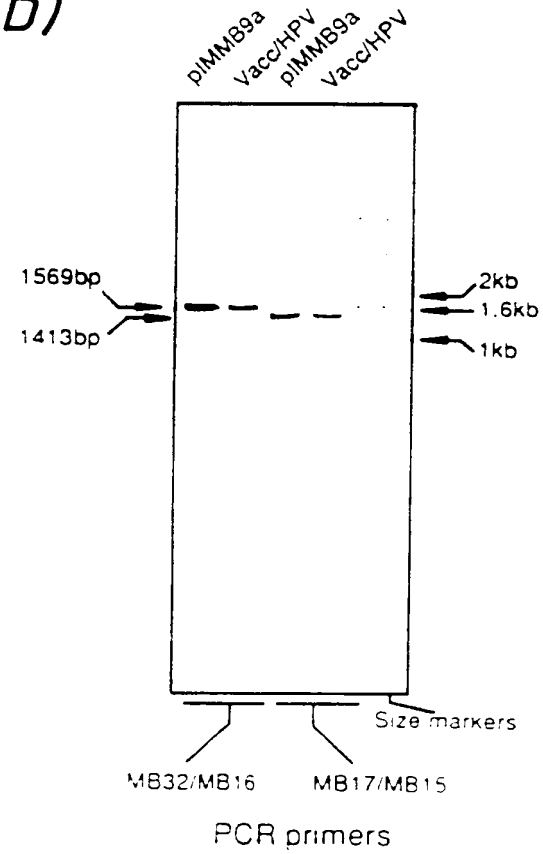


Fig. 23.

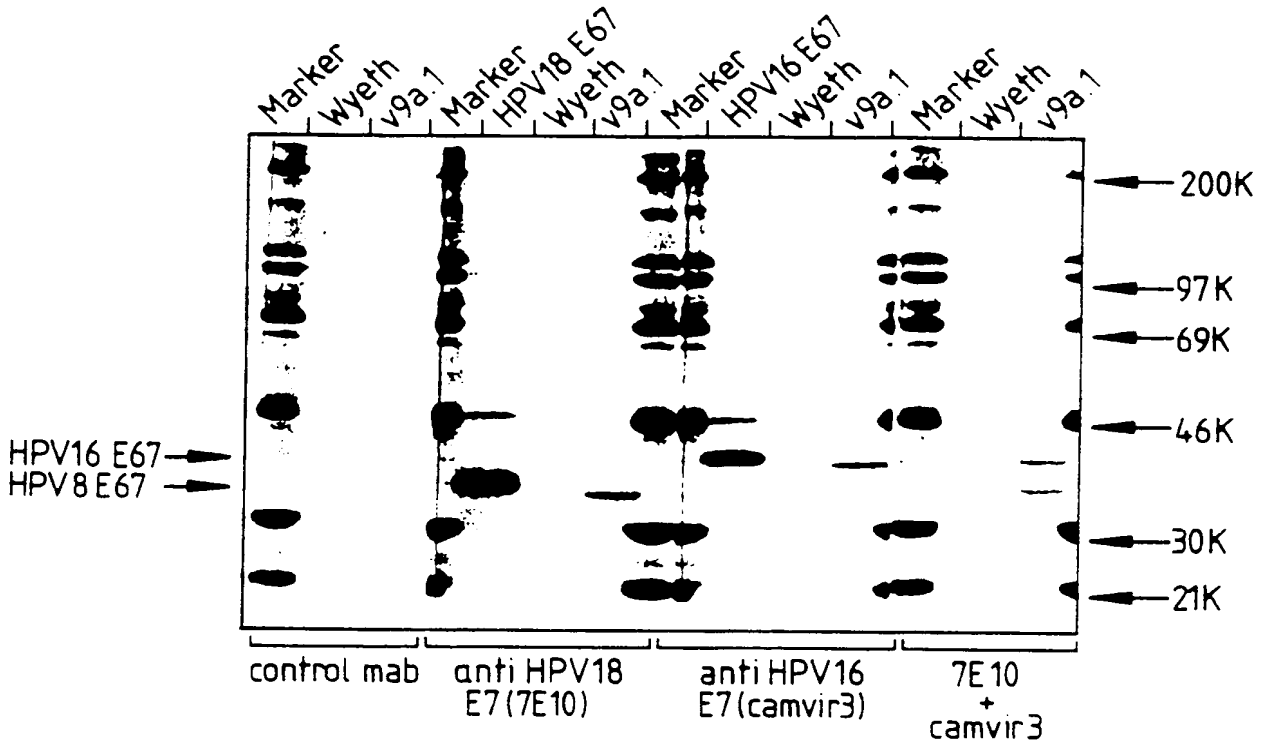
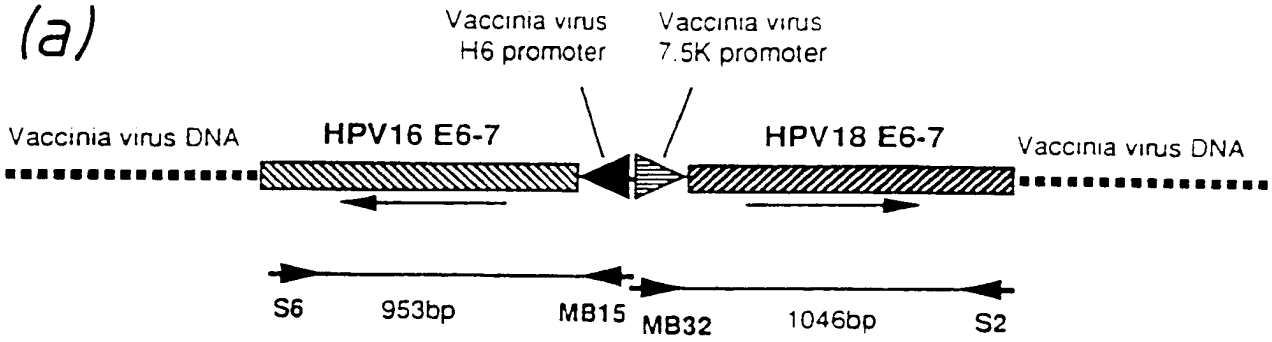

Marker
Wyeth
v9a.1
Marker
HPV18 E67
Wyeth
v9a.1
Marker
HPV16 E67
Wyeth
v9a.1
Marker
Wyeth
v9a.1
200K
97K
69K
46K
30K
21K
HPV16 E67
HPV8 E67
control mab
anti HPV18 E7 (7E10)
anti HPV16 E7 (camvir3)
7E10 + camvir3

Fig. 24.

(a)



The sequences present at the 5' end of PCR primers M15 and MB32 overlap by 6 nucleotides

(b)

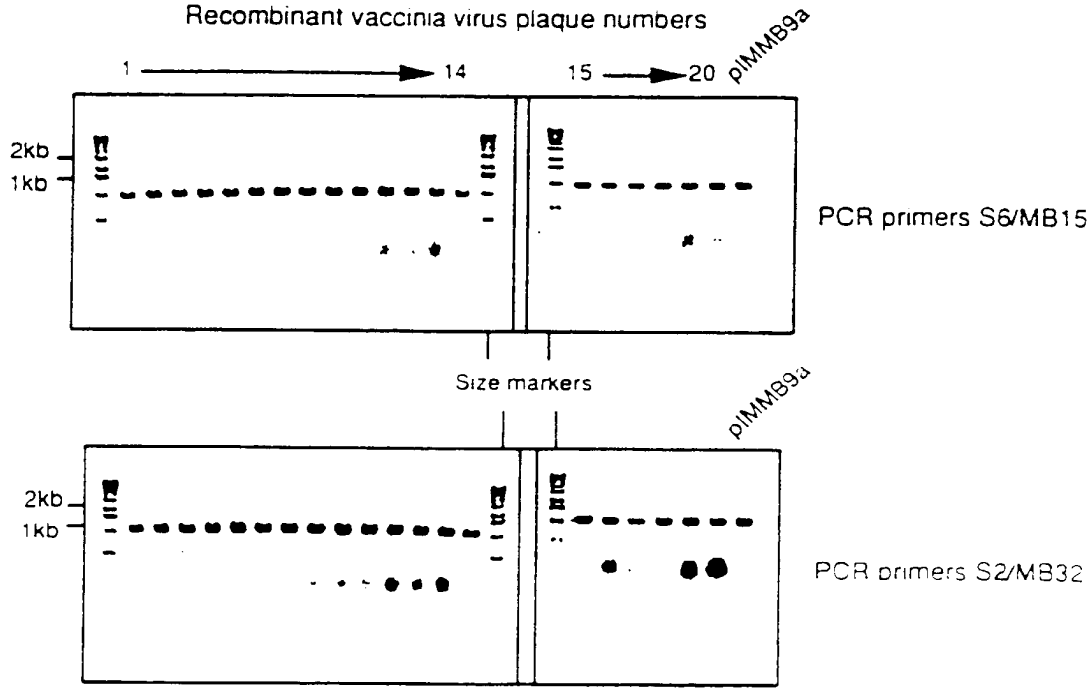


Fig. 25.

	Virus Titre (log_{10}pfu)			
	Experiment 1		*Experiment 2*	
	Wyeth	**v9a.1**	**Wyeth**	**v9a.1**
1 day	4.02	4.65	4.11	2.93
3 days	5.44	4.46	4.86	3.20
5 days	4.20	3.76	3.93	3.26

Fig. 26.

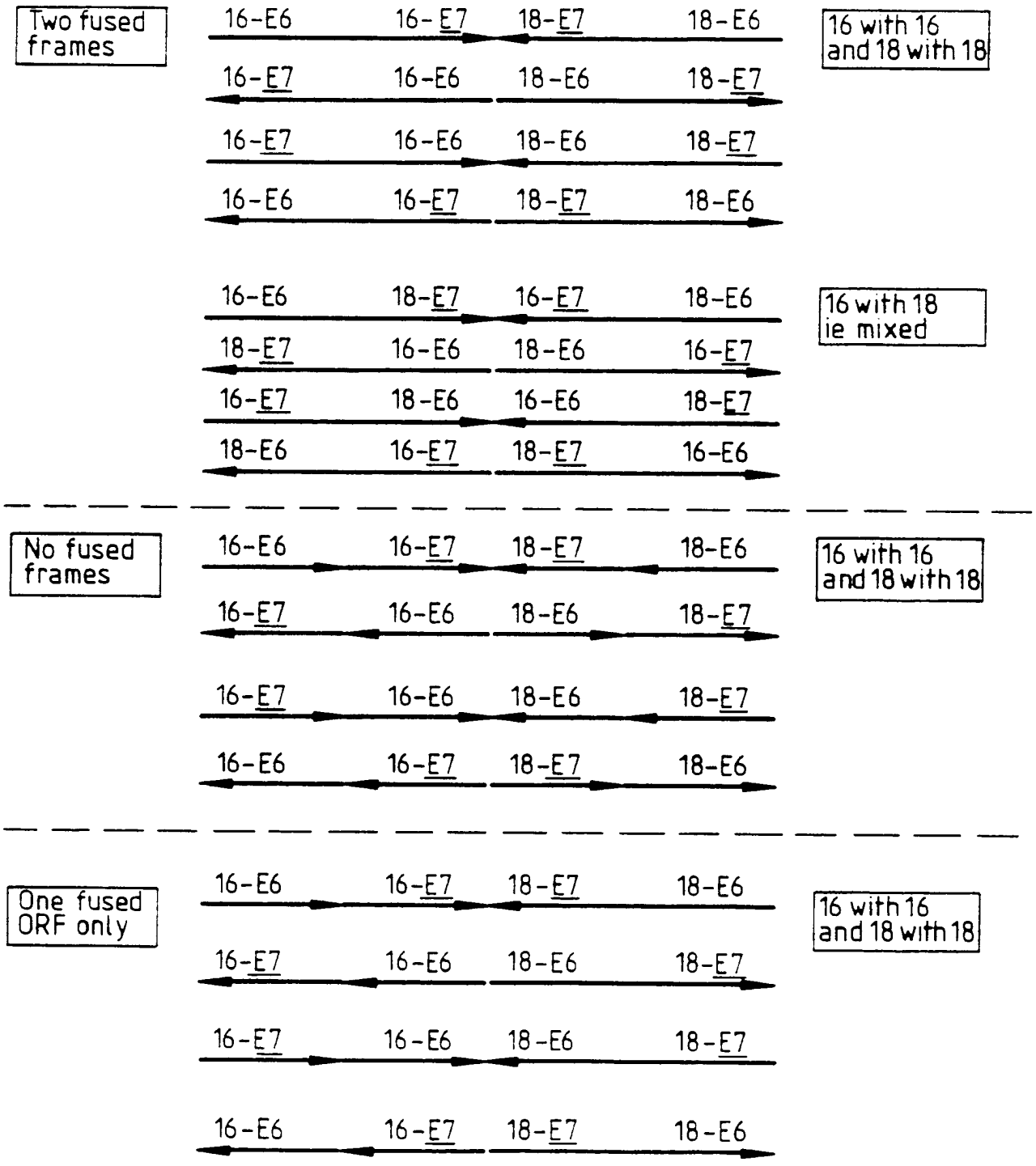
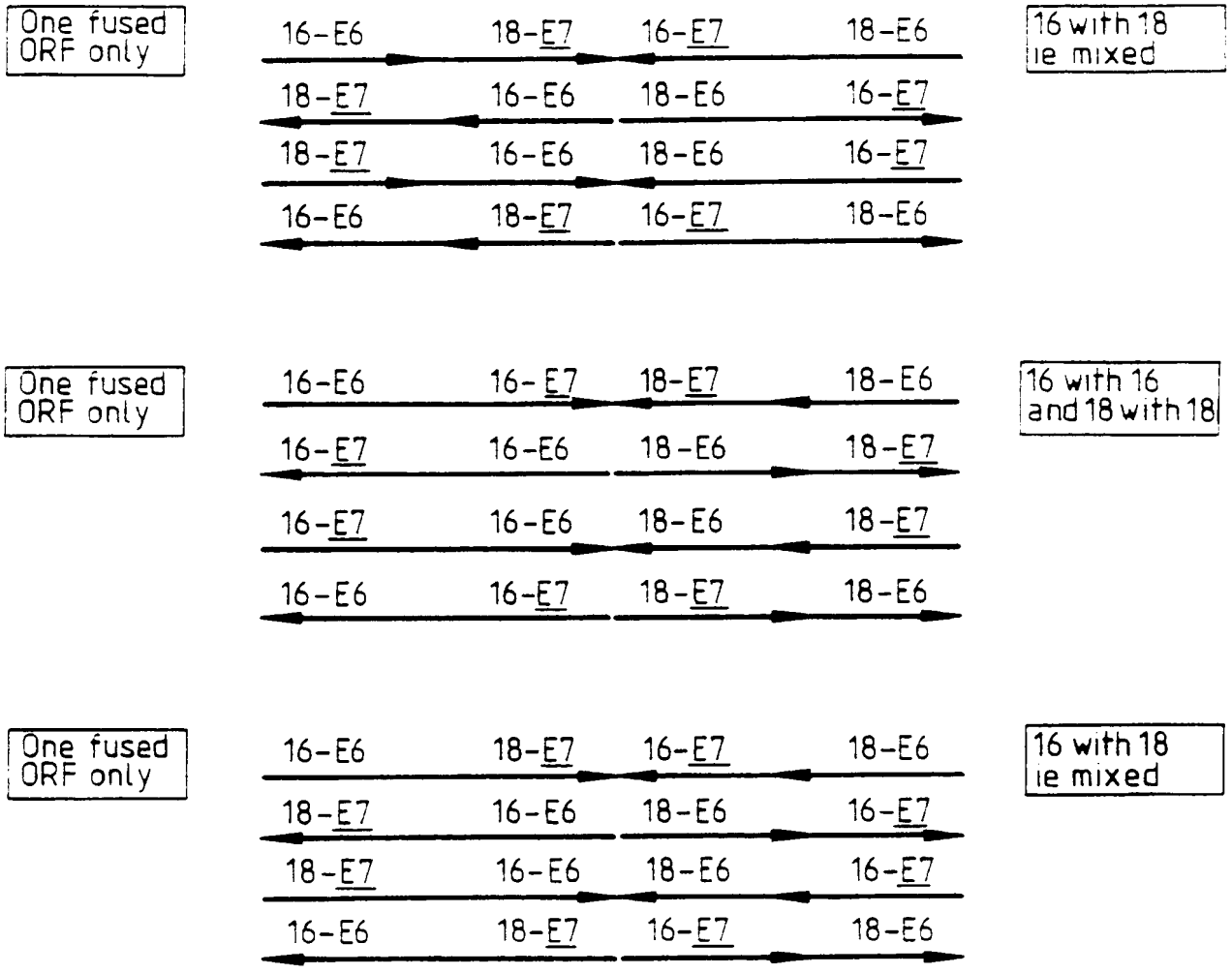

Two fused frames
16-E6 16-E7 18-E7 18-E6
16-E7 16-E6 18-E6 18-E7
16-E7 16-E6 18-E6 18-E7
16-E6 16-E7 18-E7 18-E6
16 with 16 and 18 with 18
16-E6 18-E7 16-E7 18-E6
18-E7 16-E6 18-E6 16-E7
16-E7 18-E6 16-E6 18-E7
18-E6 16-E7 18-E7 16-E6
16 with 18 ie mixed
No fused frames
16-E6 16-E7 18-E7 18-E6
16-E7 16-E6 18-E6 18-E7
16-E7 16-E6 18-E6 18-E7
16-E6 16-E7 18-E7 18-E6
16 with 16 and 18 with 18
One fused ORF only
16-E6 16-E7 18-E7 18-E6
16-E7 16-E6 18-E6 18-E7
16-E7 16-E6 18-E6 18-E7
16-E6 16-E7 18-E7 18-E6
16 with 16 and 18 with 18

Fig. 26 (Cont).


One fused ORF only
16-E6
18-E7
16-E7
18-E6
16 with 18 ie mixed
18-E7
16-E6
18-E6
16-E7
18-E7
16-E6
18-E6
16-E7
16-E6
18-E7
16-E7
18-E6
One fused ORF only
16-E6
16-E7
18-E7
18-E6
16 with 16 and 18 with 18
16-E7
16-E6
18-E6
18-E7
16-E7
16-E6
18-E6
18-E7
16-E6
16-E7
18-E7
18-E6
One fused ORF only
16-E6
18-E7
16-E7
18-E6
16 with 18 ie mixed
18-E7
16-E6
18-E6
16-E7
18-E7
16-E6
18-E6
16-E7
16-E6
18-E7
16-E7
18-E6

INTERNATIONAL SEARCH REPORT

International Application No PCT/GB 92/00424

I. CLASSIFICATION OF SUBJECT MATTER (if several classification symbols apply, indicate all)[6]

According to International Patent Classification (IPC) or to both National Classification and IPC

Int.Cl.5 C 12 N 15/86 C 12 N 15/62 C 12 N 15/37
A 61 K 39/12

II. FIELDS SEARCHED

Minimum Documentation Searched[7]

Classification System	Classification Symbols
Int.Cl.5	C 12 N C 07 K

Documentation Searched other than Minimum Documentation to the Extent that such Documents are Included in the Fields Searched[8]

III. DOCUMENTS CONSIDERED TO BE RELEVANT[9]

Category °	Citation of Document, [11] with indication, where appropriate, of the relevant passages [12]	Relevant to Claim No.[13]
X	Biochemical and Biophysical Research Communications, vol. 174, no. 1, 15 January 1991, A. NAITO et al.: "Homologous recombination in bovine papillomavirus shuttlevecter; effect of relative orientation of substrate sequences", pages 305-312, see the whole document	1
A	WO,A,9012880 (APPLIED BIOTECHNOLOGY) 1 November 1990, see the whole document	1
A	Mol. Gen. Genet., vol. 222, no. 2-3, 1990, Y. KITAMURA et al.: "Homologous recombination in a mammalian plasmid", pages 185-191, see the whole document	1
A	WO,A,9012882 (HEALTH RESEARCH) 1 Novemer 1990, see the whole document	1

-/-

° Special categories of cited documents : [10]

"A" document defining the general state of the art which is not considered to be of particular relevance

"E" earlier document but published on or after the international filing date

"L" document which may throw doubts on priority claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified)

"O" document referring to an oral disclosure, use, exhibition or other means

"P" document published prior to the international filing date but later than the priority date claimed

"T" later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention

"X" document of particular relevance; the claimed invention cannot be considered novel or cannot be considered to involve an inventive step

"Y" document of particular relevance; the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art.

"&" document member of the same patent family

IV. CERTIFICATION

Date of the Actual Completion of the International Search	Date of Mailing of this International Search Report
04-06-1992	30. 06. 92
International Searching Authority: EUROPEAN PATENT OFFICE	Signature of Authorized Officer: Natalia Weinberg

III. DOCUMENTS CONSIDERED TO BE RELEVANT (CONTINUED FROM THE SECOND SHEET)

	Category *	Citation of Document, with indication, where appropriate, of the relevant passages	Relevant to Claim No.
8	A	The EMBO Journal, vol. 7, no. 6, 1988, A. STOREY et al.: "Comparison of the in vitro transforming activities of human papillomavirus types", pages 1815-1820, see the whole document	3-7
4	A	The Journal of Biological Chemistry, vol. 265, no. 22, 5 August 1990, (Baltimore, MD, US), R.E. JONES et al.: "Identification of HPV-16 E7 peptides that are potent antagonists of E7 binding to the retinoblastoma suppressor protein", pages 12782-12785, see the whole document	14-16
4	A	WO,A,9010459 (TRANSGENE) 20 September 1990, see the whole document	2,3,18

FURTHER INFORMATIC[redacted] CONTINUED FROM THE SECOND SHEET

V. [X] OBSERVATION WHERE CERTAIN CLAIMS WERE FOUND UNSEARCHABLE [1]

This International search report has not been established in respect of certain claims under Article 17(2)(a) for the following reasons:

1 [X] Claim numbers 24 partially because they relate to subject matter not required to be searched by this Authority, namely:

Although claim 24 as far as cncerning a method in vivo is directed to a method of treatman of the human/animal body the search has been carried out and based on the alleged effects of compound/composition

2 [ ] Claim numbers because they relate to parts of the international application that do not comply with the prescribed requirements to such an extent that no meaningful international search can be carried out, specifically:

3 [ ] Claim numbers because they are dependent claims and are not drafted in accordance with the second and third sentences of PCT Rule 6.4(a).

VI. [ ] OBSERVATIONS WHERE UNITY OF INVENTION IS LACKING [2]

This International Searching Authority found multiple inventions in this international application as follows:

1 [ ] As all required additional search fees were timely paid by the applicant, this international search report covers all searchable claims of the international application

2 [ ] As only some of the required additional search fees were timely paid by the applicant, this international search report covers only those claims of the international application for which fees were paid, specifically claims:

3 [ ] No required additional search fees were timely paid by the applicant. Consequently, this international search report is restricted to the invention first mentioned in the claims; it is covered by claim numbers:

4 [ ] As all searchable claims could be searched without effort justifying an additional fee, the International Searching Authority did not invite payment of any additional fee.

Remark on Protest

[ ] The additional search fees were accompanied by applicant's protest.

[ ] No protest accompanied the payment of additional search fees.

ANNEX TO THE INTERNATIONAL SEARCH REPORT ON INTERNATIONAL PATENT APPLICATION NO. GB 9200424 SA 57334

This annex lists the patent family members relating to the patent documents cited in the above-mentioned international search report. The members are as contained in the European Patent Office EDP file on 23/06/92
The European Patent Office is in no way liable for these particulars which are merely given for the purpose of information.

Patent document cited in search report	Publication date	Patent family member(s)	Publication date
WO-A- 9012880	01-11-90	EP-A- 0469089	05-02-92
WO-A- 9012882	01-11-90	AU-A- 5552090	16-11-90
		CA-A- 2014465	17-10-90
		DE-T- 4090565	14-05-92
		FR-A- 2647808	07-12-90
		GB-A- 2246784	12-02-92
		NL-A- 9020677	03-02-92
WO-A- 9010459	20-09-90	FR-A- 2643817	07-09-90
		EP-A- 0462187	27-12-91